Social Sci. - Periodical

Social Science Library

BF 1 J86

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California Los Angeles*
Neal E. Miller, *Yale University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Studies in Incidental Learning: II. The Effects of Association Value and of the Method of Testing: L. POSTMAN, P. A. ADAMS, AND L. W. PHILLIPS 1

Transposition in the Feebleminded: H. W. STEVENSON AND I. ISCOE 11

Probability Learning in a Problem-Solving Situation: J. J. GOODNOW AND L. POSTMAN .. 16

The Relation of Conditioned Discrimination to the MMPI *Pd* Personality Variable: A. B. WARREN AND D. A. GRANT 23

Learning Sets from Minimum Stimuli: A. J. RIOPELLE 28

Distribution of Practice and the Temporal Decay of Response-Produced Stimuli: E. Z. ROTHKOPF 33

Learning and Performance on a Key-Pressing Task as Function of the Degree of Spatial Stimulus-Response Correspondence: R. E. MORIN AND D. A. GRANT 39

Rotary Pursuit Performance Under Alternate Conditions of Distributed and Massed Practice: M. R. DENNY, N. FRISBEY, AND J. WEAVER, JR. 48

A Comparison of Pursuit and Compensatory Tracking Under Conditions of Aiding and No Aiding: R. CHERNIKOFF, H. P. BIRMINGHAM, AND F. V. TAYLOR 55

Mitigation of Work Decrement: G. T. HAUTY AND R. B. PAYNE 60

Variables Affecting the Angular Displacement Threshold of Simulated Auditory Movement: R. C. WILCOTT 68

American Psychological Association

Vol. 49 No. 1 January 1955

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1955 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Box 1557, Lackland Air Force Base, San Antonio, Texas until March 19, 1955. From March 19 to June 1 address manuscripts to the editor at Department of Psychology, Stanford University, Stanford, California. Manuscripts sent after June 1 should be addressed to San Antonio again. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

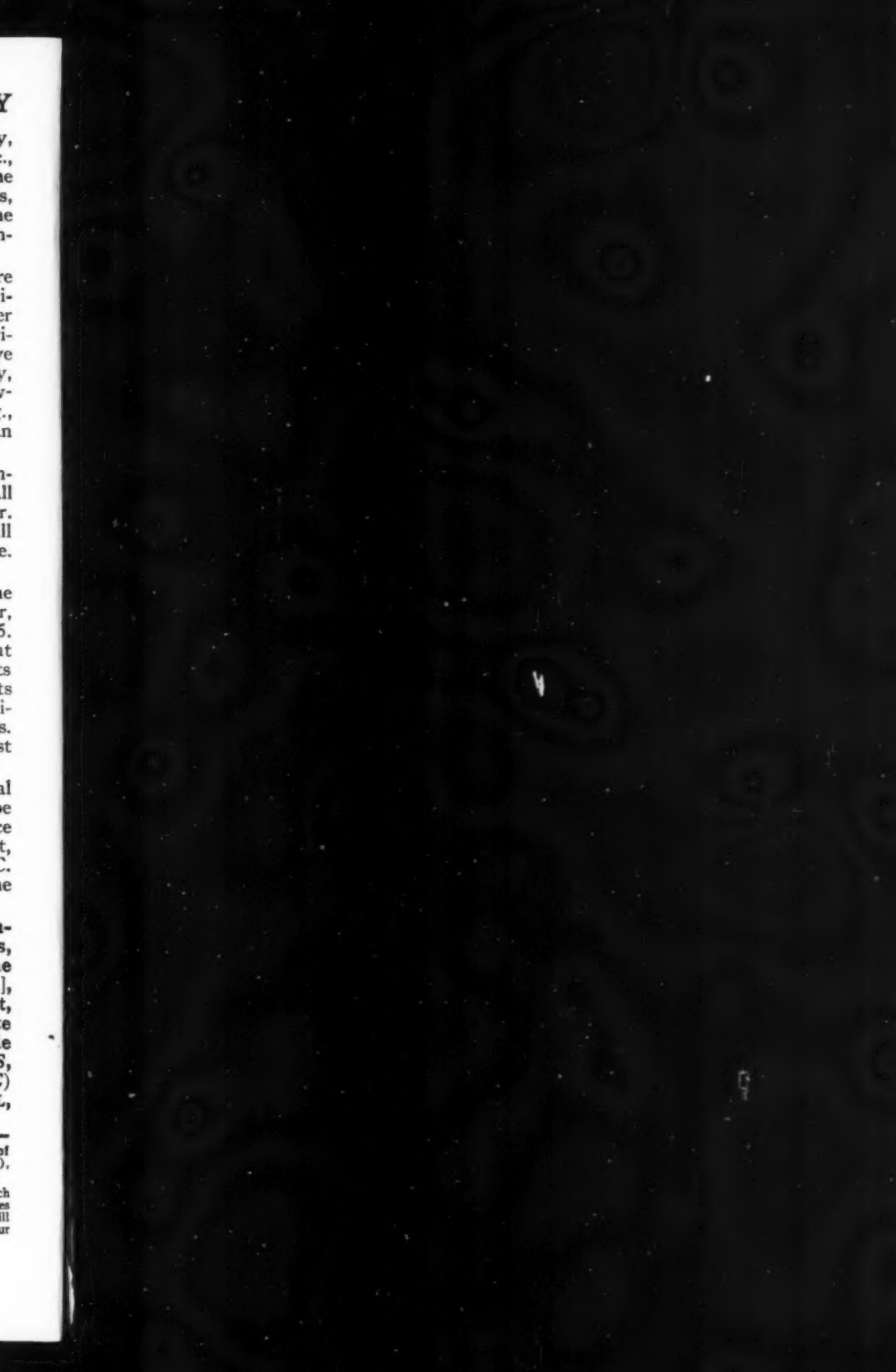
To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright, 1955, by the American Psychological Association, Inc.



Journal of
Experimental Psychology

VOL. 49, NO. 1 JANUARY, 1955

STUDIES IN INCIDENTAL LEARNING: II. THE EFFECTS OF ASSOCIATION VALUE AND OF THE METHOD OF TESTING[1]

LEO POSTMAN, PAULINE AUSTIN ADAMS, AND LAURA W. PHILLIPS

University of California

Incidental learning is a selective process. Typically, the incidental learner can recall only a fraction of the materials to which he has been exposed. There appear to be two major approaches to the analysis of the conditions governing the selectivity of incidental learning. One may seek the determinants of selectivity primarily in (*a*) the perceptual characteristics of the stimulus materials, or (*b*) the response habits with which the incidental learner enters the experimental situation. A previous study (6), which failed to find a difference in incidental memory for crowded and isolated items, cast serious doubts on the adequacy of the perceptual approach. The present investigation focuses on *Ss*' response habits as determiners of selective retention under incidental conditions.

We shall assume that the amount of incidental as well as intentional learning depends to an important degree on the differential responses which *S* makes to the individual members of the stimulus sequence. Differential responses to a series of items may include (*a*) classifying or categorizing responses such as naming or labeling; (*b*) other responses elicited by the items by virtue of stimulus generalization; (*c*) responses serving as associative links among the members of the series. Such responses are associated with the context of the experimental situation (cf. 4). When *S*'s recall is tested, the rearousal of these differential responses mediates the recall performance.

It is our hypothesis that there are major differences between intentional and incidental learners in the frequency and intensity of differential responses to the stimulus items. Differential responses may be regarded as habits attached, largely through stimulus generalization, to the individual items of the series. The actual performance of such responses will be a function of both the strength of the associative habits and *S*'s motivation. Intentional learners, under instructions to learn the series, are more highly motivated to respond to each of the items. Assuming equal initial strengths of differential responses for both types of learners, the

[1] This research was facilitated by a grant from the Behavioral Sciences Division of the Ford Foundation.

threshold of response will be exceeded more readily for the intentional learners. Intentional learners may, therefore, be expected to give more frequent differential responses to the stimulus items. A further difference between the two types of learners may be sought in the intensity ("amplitude") of whatever differential responses are made, i.e., responses of the intentional learner may be more explicitly verbalized, of longer duration, etc. Finally, it is probable that intentional learners rehearse their differential responses more than do the incidental learners during exposure to the stimulus materials. In short, we assume that there is no sharp qualitative break between intentional and incidental learning. The difference is a matter of degree, particularly as regards the readiness to respond differentially to the stimulus items.

This analysis implies that the difference in the amount recalled by intentional and incidental learners should be a function of the association value of the stimulus items. The association value of an item may be regarded as an index of the readiness with which it elicits differential responses. Stimuli of high association value are likely to evoke effective differential responses from both intentional and incidental learners, and both groups should learn such items well. As items decrease in association value, they are likely to evoke less effective differential responses from incidental than intentional learners. Consequently, the lower the association value of the stimuli, the greater should be the difference between the amount retained by intentional and incidental learners.

If the learning materials consist of items differing in association value, we should expect systematic differences between the distributions of associations formed under intentional and incidental conditions. Incidental learning will be selective: some strong associations will be formed, notably for the relatively familiar items, while the associations for the remaining less familiar items will be weak. The associations formed by intentional learners will be more continuously graded in strength. In addition to strong associations for relatively familiar items, intentional learners will form more associations of intermediate strength and fewer weak associations than the incidental learners. Median associative strength will, therefore, be higher under intentional than incidental conditions.

This picture of the distribution of associations suggests that the difference in performance between intentional and incidental learners should vary systematically with the difficulty of the retention test. We should expect neither a very easy nor a very difficult retention test to discriminate effectively between the intentional and incidental conditions. If a test is relatively easy, only a low degree of associative strength will be required for correct performance. The weak associations formed by incidental learners will then be as adequate as the stronger ones formed by intentional learners. If the test is relatively difficult, only associations of high strength will result in correct performance. Such a test would tap only the core of strongly learned items which we have assumed to be common to the two groups of learners. A test of intermediate difficulty, i.e., a test requiring an intermediate degree of associative strength for correct performance, should discriminate best between intentional and incidental learners. This prediction is based on the assumption that median associative strength is higher for intentional than incidental learners.

Let us assume that among possible tests of retention a test of free recall occupies a position of intermediate difficulty. It follows from the above analysis that a test of recognition, which is sensitive to low degrees of learning and hence relatively easy, should yield smaller differences between intentional and incidental learners than a test of free recall. In addition to changing from recall to recognition, one may also vary the difficulty of the retention test by manipulation of the stimulus context at the time of recall (**3,** pp. 501–505). The associative context in which retention is tested may be such as to facilitate or inhibit recall. As compared with a standard test of free recall, a facilitating context would define a relatively easy test, and an inhibitory context would define a relatively difficult test of retention. The introduction of either a facilitating or an inhibitory context should level the difference between intentional and incidental learners.

The present study presents two experiments testing the implications of our analysis. In Exp. I, items of different association value were used, and the performance of intentional and incidental learners was tested both by free recall and recognition. Experiment II is concerned with the effects of facilitating and inhibitory contexts on the recall performance of the two kinds of learners.

Experiment I

Method

Conditions of learning.—The procedure used to create conditions of intentional and incidental learning has been described in detail in a previous communication (**6**). Two individuals were used at a time, one of whom served as *E* and the other as *S*. The *E* and the *S* were seated in front of a memory drum on which a series of nonsense syllables was exposed. As each syllable appeared in the window of the drum, *E* spelled it aloud, letter by letter. The *S* repeated the syllable as soon as it had been read by *E*. The *S* was instructed to learn as many items as possible though not necessarily in the order in which they were presented. No learning instructions were given to *E*. The *S* was an intentional learner and *E* an incidental learner.

Materials.—The stimulus list consisted of 20 nonsense syllables differing widely in association value. There were five syllables each of 0, 33, 66, and 100% association value by Glaze's norms (**1**). The series was divided into five successive blocks of four syllables each. Each block contained one syllable of each association value so that the various association values were evenly distributed over different parts of the list.

The syllables were presented on a Hull-type memory drum at the rate of one every 4 sec. The list was presented only once. It was hoped that restriction to one exposure would minimize the probability of *E*'s adopting a set to learn. A group of four syllables was used to practice the experimental procedure, after which the experimental list was presented.

Retention tests.—For purposes of measuring retention, both *S*s and *E*s were divided into a Recall Group and a Recognition Group. Each group was given two tests, one immediately after presentation of the stimulus list, and the other 20 min. later. The interval between tests was filled with a reading task which was presented as a test of reading speed and comprehension. Selected passages from Plato's *Dialogues* were used for the reading task.

On both recall tests, *S*s and *E*s were instructed to write down as many syllables as they could remember, regardless of the order of original presentation. The recognition test consisted of 20 multiple-choice items. Each multiple-choice item consisted of four syllables, one of which was taken from the learning series. None of the incorrect alternatives had more than one letter in common with any of the correct syllables. The *S*s and *E*s were forced to guess if necessary, i.e., they were required to choose one of the alternatives in each of the multiple-choice items. All tests had a time limit of 5 min.

Postexperimental inquiry.—Both *S*s and *E*s were given a standard interview at the end of the experimental session. The *E*s were asked whether (*a*) they had expected the first retention test; (*b*) they had attempted to learn any of the syllables; (*c*) which of the syllables, if any, had caught their attention during the presentation of the series; (*d*) they had expected the second retention test; (*e*) they had rehearsed any of the syllables during the interval between retention tests. The *S*s were asked only questions *c*, *d*, and *e*. The *E*s who had attempted to memorize the syllables during presentation of the series or who had expected the first retention test were discarded.

TABLE 1

MEAN NUMBER OF CORRECT RECALLS AND RECOGNITIONS BY *Ss* AND *Es* (EXP. I)

Group	Recall				Recognition			
	Test I		Test II		Test I		Test II	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Ss	3.63	1.56	3.20	1.87	12.40	2.90	11.15	3.12
Es	3.03	1.82	2.70	1.79	12.50	2.18	11.55	2.97

Subjects.—There were 30 *Ss* and 30 *Es* in the Recall Groups and 20 pairs in the Recognition Groups. All *Ss* and *Es* were undergraduate students at the University of California. They did not know the purpose of the experiment.

Results

Amount of retention.—Table 1 presents the results of the tests of recall and recognition. On Test I, the over-all recall performance of the *Ss* is higher than that of the *Es*, but the difference is not statistically significant ($t = 1.36$, $p > .05$). Both groups show a significant drop in recall on Test II. Comparison of Test I and Test II yields a t of 2.87 ($p < .01$) for *Ss*, and a t of 2.36 ($p = .02$) for *Es*. Analysis of covariance, which takes the original level of learning into account, shows that there is no significant difference between intentional and incidental learners in the amount of retention loss ($F = .12$).

With the amount of practice held constant, the recognition tests yield considerably higher retention scores than do the tests of free recall. The method of recognition is more sensitive than the method of free recall, i.e., the minimum associative strength required for correct performance is lower for recognition than for recall. The recognition scores of *Ss* and *Es* are virtually identical on both Test I and Test II. The difference in favor of intentional learners found in free recall disappears completely when a test of recognition is used.

Both groups show drops in their recognition scores on Test II. For *Ss*, comparison of Test I and Test II yields a t of 2.45 ($p = .03$); for *Es*, $t = 1.94$ ($p = .07$). Although the

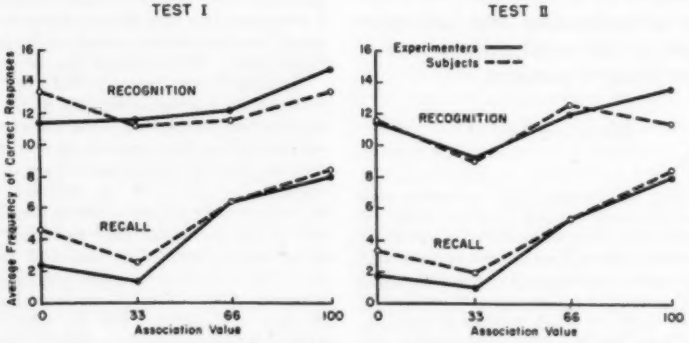


FIG. 1. Average frequency of correct recalls and recognitions by *Ss* and *Es* as a function of association value (Exp. I).

percentage of loss is somewhat smaller for *Es* than for *Ss*, analysis of covariance again fails to reveal a significant difference in retention loss between the two groups ($F = .42$).

Recall and recognition as a function of association value.—Figure 1 shows the average frequency of recall and recognition for syllables of different association value. Each point represents an average for five syllables distributed over different serial positions in the list.

Association value has a differential effect on the recall performance of intentional and incidental learners. Intentional learners recall more syllables of low association value than do incidental learners. The recall scores of the two groups are virtually identical for syllables of high association value.

To evaluate the significance of the relationship between condition of learning and association value, the median test (**5**, p. 394) was applied to the results of Test I. Association values were dichotomized into high (66% and 100%) and low (0% and 33%). For each *S* and *E*, the number of high-association-value syllables recalled was expressed as a proportion of the total recall score, and the median proportion was computed for the combined groups. The proportions were then divided into those greater than and equal to the median, and those falling below the median. By chance, half the members of each group should fall above the common median and half below. The null hypothesis can be rejected at a high level of confidence ($\chi^2 = 9.88$, $df = 1$, $p < .01$).

The results of Test II are substantially the same as those of Test I. It is interesting to note that both intentional and incidental learners show relatively greater retention loss for items of low association value than for items of high association value. There is no loss at all for syllables of 100% association value.

It is now apparent why the *over-all* difference between *Ss* and *Es* fails to reach statistical significance. Comparison of the total recall scores does not take account of the selectivity of incidental learning. Intention to learn is more essential for the recall of relatively unfamiliar than of relatively familiar items.

The differential effects of association value are much less pronounced for recognition than they are for recall. For *Ss*, there is little, if any, systematic relationship between association value and recognition. In the case of *Es*, recognition tends to increase as a function of association value on both tests. However, the degree of divergence between intentional and incidental learners is small. Application of the median test shows that *Es* do not surpass *Ss* significantly in the selective recognition of items of high association value ($\chi^2 = 3.14$, $df = 1$, $p > .05$).

Postexperimental inquiry.—The second retention test was expected by five *Ss* and three *Es* in the Recall Groups and by seven *Ss* and three *Es* in the Recognition Groups. One *S* and one *E* in the Recall Groups, but none of the *Ss* and *Es* in the Recognition Groups, reported rehearsing the material between retention tests. Most *Ss* and *Es* mentioned at least one item which had caught their special attention during the presentation of the stimulus list. As in our previous study (**6**), virtually all items singled out by *Ss* and *Es* were noticed because of their similarity to a meaningful word or symbol.

Experiment II

The results of the recognition test have shown that a test easier than free recall reduces the difference be-

tween intentional and incidental learners. It remains necessary to demonstrate that a test more difficult, as well as a test easier, than free recall results in a convergence of the two groups of learners. In addition, it is desirable to test the implications of our analysis while holding the nature of the memory performance constant. Free recall and recognition require different types of responses of the *S*. The question remains open as to whether variations in the difficulty of the retention test will have the predicted effects when only one type of performance, e.g., active reproduction, is used. In Exp. II the difficulty of the retention test was varied by manipulation of the stimulating context at the time of recall.

Method

Conditions of learning.—The learning materials consisted of 30 adjectives, taken from the list published by Haagen (**2**). An "orienting task" was used in order to expose both intentional and incidental learners to the stimulus materials under identical conditions. Both groups of learners were required to rate each of the adjectives with respect to the frequency with which they used it in their daily speech. A 7-point rating scale was used, with 1 standing for extremely infrequent, and 7 for extremely frequent, with the other numbers indicating intermediate degrees of frequency. The adjectives, typed in capital letters, were presented in a booklet. Each adjective appeared on a separate page of the booklet. The pages were numbered consecutively from 1 through 30. When *E* announced the number 1, *Ss* turned to the first page of the booklet and made their rating by circling the appropriate number on a 7-point scale printed underneath the word. The same procedure was followed throughout the presentation of the list. The successive stimulus words were presented at the rate of one every 10 sec. A large model of the rating scale, with each of the steps labeled, was kept in full sight of the *Ss* throughout the rating period. The list was presented once.

There were two forms of the booklet, each with a different random order of the stimulus words. The two orders were used equally often under all conditions of the experiment.

Intentional learners were informed that their memory for the words in the booklet would be tested later. They were asked to remember as many of the items as possible, though not necessarily in the order in which they were presented. No learning instructions were given to the incidental learners.

Retention tests.—Both intentional and incidental learners were divided into three groups for purposes of the retention test: (*a*) Free Recall Groups of 32 *Ss* each, (*b*) Facilitating Context Groups of 16 *Ss* each, and (*c*) Inhibitory Context Groups of 16 *Ss* each. Members of the Free Recall Groups were given 5 min. in which to write down as many of the stimulus words as they could remember, regardless of the original order of presentation. They were encouraged to put down guesses as well as items of which they felt certain. The two Context Groups were provided with a series of cue words to which *Ss* were instructed to respond with words from the original list. For the Facilitating Context Groups, the cue words had close associative connections with the original stimulus words. These cues were, therefore, expected to facilitate the evocation of correct responses. For the Inhibitory Context Groups, the cue words had only remote associative connections with the original stimulus words and were expected to elicit responses which would compete with the evocation of the correct items. Since the stimulus words had been taken from Haagen's list, facilitating and inhibitory cues were chosen on the basis of Haagen's "A" scales (**2**) which measure the "closeness of associative connection" between standard stimulus words and sets of comparison words. The range of scale separations in Haagen's tables is from .5 to 6.0. For the Facilitating Context Groups, the scale separations between the cue words and learning words were less than, or equal to, 1.5. The average scale separation was 1.2, with an *SD* of .2. For the Inhibitory Context Groups, the scale separations were equal to, or larger than, 3.0. The average scale separation was 3.8, with an *SD* of .6. Each cue word appeared on a separate page of the test booklet in which *Ss* recorded their responses. The cue words were presented at the rate of one every 20 sec. For both lists of cue words there were two different random orders of presentation which were used equally often. The *Ss* were encouraged to guess if they were not sure about a response word.

Postexperimental inquiry.—Following the retention test, all incidental learners were asked whether they had expected to be tested. There were no such cases, indicating that the orienting task had been quite effective in disguising the purpose of the experiment.

Association Control Groups.—In order to evaluate the effects of the cue words on retention, it was necessary to determine how many correct

responses would be elicited by the cue words independently of previous learning. There were two Association Control Groups of 20 *Ss* each. One of these control groups was presented with the same cue words as the Facilitating Context Groups, the other group responded to the same cue words as the Inhibitory Context Groups. The control *Ss* were instructed to write down next to each cue word one adjective which they considered similar in meaning to the word. The cue words were again presented at the rate of one every 20 sec.

Subjects.—The *Ss* were undergraduate students at the University of California. They did not know the purpose of the experiment and were assigned to the experimental conditions at random.

Results

Retention as a function of associative context.—Table 2 shows the average number of items correctly reproduced by each of the experimental groups. In the case of the Context Groups all correct reproductions are included, regardless of whether or not the responses were made to the appropriate cues. The amount recalled varies widely as a function of the method of testing. The rank order of conditions is the same for intentional and incidental learners. The Facilitating Context Groups recall the most, the Inhibitory Context Groups the least, and the Free Recall Groups occupy an intermediate position. For both intentional and incidental learners, *t* tests show all differences between methods of testing to be highly significant. The values of p range from .015 to $< .0001$.

Under both conditions of learning, the Context Groups performed significantly better than the corresponding Association Control Groups. All tests of significance comparing Context Groups with the appropriate Association Control Groups yield values of *t* significant at the better than 1% level of confidence. All Context Groups show a significant amount of learning but the level of their performance varies markedly with the nature of the cue words.

The size and significance of the difference between intentional and incidental learners depend on the method of testing, i.e., on the stimulating conditions at the time of recall. Only the method of free recall yields a significant difference in favor of intentional learners. For both the Facilitating Context Groups and the Inhibitory Context Groups, the performance of intentional and incidental learners is virtually identical.

Free recall as a function of familiarity.—Experiment I has shown that the difference between intentional and incidental learners depends on the familiarity of the stimulus items. As the familiarity of the items increases, the performance of the two groups converges. The results of the free recall test in the present study provide further support for this relationship.

TABLE 2

MEAN NUMBER OF ITEMS CORRECTLY REPRODUCED BY THE DIFFERENT GROUPS IN EXP. II AND SIGNIFICANCE OF DIFFERENCES BETWEEN INTENTIONAL AND INCIDENTAL LEARNERS

Condition	Association Control		Intentional		Incidental		Int. vs. Incidental Learners
	Mean	*SD*	Mean	*SD*	Mean	*SD*	*t*
Facilitating context	9.10	3.05	17.62	4.20	18.00	4.58	0.24
Free recall	—	—	12.62	3.73	9.41	2.58	4.31*
Inhibitory context	1.05	0.80	7.19	2.13	7.38	2.44	0.23

* $p < .01$.

During the original exposure to the learning list, all *Ss* rated their frequency of usage of the stimulus items on a 7-point scale. The distributions of ratings were almost identical for intentional and incidental learners. The mean rating for intentional learners was 3.91, with an *SD* of 1.58; the mean rating for incidental learners was 3.88, with an *SD* of 1.48. For both groups, the ratings showed a marked central tendency, so that the distributions were approximately normal in shape. Figure 2 shows for the two groups of learners the percentage of items correctly reproduced on the free recall test as a function of familiarity rating. Although neither the reliability nor the validity of the ratings is known, we may assume that they provide at least a crude measure of the relative frequency of usage. As Fig. 2 shows, intentional learners surpass incidental learners in the reproduction of relatively unfamiliar words, whereas there is only a small difference between the two groups in the recall of more familiar items.

Relatively unfamiliar items were recalled better than items of medium familiarity, particularly by intentional learners. It is possible that items of low and high familiarity were subject to less intraserial interference than were items of medium familiarity. Medium ratings were given more frequently than high or low ratings. To the extent that items given similar ratings were grouped and perceived as similar by the *Ss*, there was more opportunity for intraserial interference among items of medium familiarity than among items of high or low familiarity.

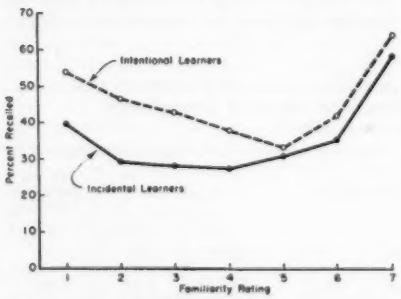


FIG. 2. Percentages of items correctly reproduced on the free recall test in Exp. II as a function of familiarity rating.

If the sag in the familiarity function is, indeed, due to intraserial interference, it is not surprising that intentional learners show the effect to a more marked degree than do incidental learners. As we have suggested in an earlier paper, incidental learners form not only fewer and/or weaker associations than intentional learners but also fewer incorrect ones. Hence the variables influencing degree of intraserial interference influence the recall performance of incidental learners less than that of intentional learners. The finding that "isolated" items are recalled better than crowded ones by intentional learners but not by incidental learners was explained along similar lines (6).

Facilitating and inhibitory effects of context.—We have suggested that the facilitating and inhibitory cues presented to the Context Groups render ineffective the difference in associative strength between intentional and incidental learners. These "leveling" effects of the contextual cues are confirmed by an analysis of the errors of the Contextual Groups. The responses of *Ss* in the four Context Groups were divided into three categories: (*a*) correctly placed recalls, i.e., correct reproductions given in response to the appropriate cues, (*b*) misplaced recalls, i.e., items from the learning list given in response to other than the appropriate cues, and (*c*) importations, i.e., items which did not appear in the original learning list. Table 3 shows the percentages of responses falling into each of these

TABLE 3

DISTRIBUTIONS OF RESPONSES GIVEN BY CONTEXT GROUPS UNDER INTENTIONAL AND INCIDENTAL CONDITIONS

(Entries are percentages of total numbers of responses)

Category of Response	Intentional		Incidental	
	Facilitating Context	Inhibitory Context	Facilitating Context	Inhibitory Context
Correctly placed	72.58	21.93	75.60	23.17
Misplaced	8.36	36.40	8.22	37.40
Importations	19.06	41.67	16.18	39.43

three categories. Facilitating and inhibitory cues produce quite different distributions of responses. In the presence of facilitating cues, the large majority of correct responses is given to the appropriate cues. The ratio of correctly placed to misplaced responses is of the order of 9:1. In the presence of inhibitory cues, the percentage of correctly placed responses is considerably smaller than the percentage of misplaced ones. The ratio of misplaced to correctly placed responses is of the order of 1.7:1. Finally, the percentages of importations are much higher for the Inhibitory Context Groups than for the Facilitating Context Groups. Most of the importations clearly consisted of words which have stronger associative connections with the cue words than do the correct responses.

The variations in the distributions of responses clearly bring out the differences between the effects of the two sets of cues. At the same time we note, however, that the distributions for intentional and incidental learners are very much alike. They differ nowhere by more than a few percentage points. The differences in associative strength between the two groups of learners, which produce significant variations on the test of free recall, are counteracted by the facilitating and inhibitory effects of the contextual cues.

DISCUSSION

The experiments have shown that the difference between intentional and incidental learners is a function of both the association value of the stimuli and the method of measuring retention. These findings agree with the view that (*a*) degree of learning is a function of the differential responses made to the stimulus items and (*b*) intention to learn enhances the effectiveness of such responses.

The higher the association value of stimulus items, the less difference there is between intentional and incidental learners. Items of high association value are recalled equally well by intentional and incidental learners because strong differential responses presumably are evoked by these items independently of instructions to learn. Such responses are not as readily evoked by items of low association value. In the case of the latter items, instructions to learn serve to increase the frequency and intensity of differential responses to a significant degree. In general, the less the initial strength of the responses mediating acquisition, the more critical does the presence of a learning set become.

The power of a retention test for discriminating between intentional and incidental learners depends on the associative strength required for correct performance. Neither a very easy nor a very difficult test effectively discriminates between the two conditions of learning. Tests of intermediate difficulty maximize the observable effects of instructions to learn. The differences

obtained by the method of free recall disappear when retention is tested by recognition or by recall in a facilitating or inhibitory context. The difficulty of a retention test, and hence its power for discriminating between intentional and incidental learners, appears to vary with the degree to which the test reinstates the original learning stimuli. The original learning items actually appear on a test of recognition; a facilitating context at recall is designed to further the reinstatement of the stimulus items. Hence, performance on both these tests requires a low degree of associative strength which can be met or exceeded with equal frequency by intentional and incidental learners. An inhibitory context at recall not only fails to reinstate the original items but also makes their reproduction difficult by evoking strong competing responses. For both groups, only a few selectively learned items exceed the threshold of recall, and again there is no difference between intentional and incidental learners. A standard test of free recall also fails to reinstate the original items but does not actively interfere with their reproduction. An intermediate degree of associative strength is, therefore, required for correct performance on the free recall test; it is only under these conditions that the difference in strength between intentional and incidental learning leads to observable differences in recall.

Summary

Differences in the amounts retained by intentional and incidental learners were compared as a function of (*a*) the association value of the stimuli and (*b*) the method of measuring retention. In Exp. I, a list of nonsense syllables covering a wide range of association values was used. Retention was measured by free recall and recognition. Intentional learners surpassed incidental learners in the recall of items of low association value but not of items of high association value. There was no significant difference in the recognition performance of the two groups.

In Exp. II, recall for a series of adjectives was measured under different stimulating conditions at the time of the retention test. Intentional and incidental learners were divided into three groups: (*a*) Free Recall Groups, (*b*) Facilitating Context Groups which were presented with cue words having strong associative connections with the learning items, and (*c*) Inhibitory Context Groups presented with cue words designed to elicit responses which would compete with the correct items. Under both conditions of learning recall was highest for the Facilitating Context Groups, lowest for the Inhibitory Context Groups, with the Free Recall Groups occupying an intermediate position. Intentional learners surpassed incidental learners on the test of free recall but under neither of the context conditions.

A major difference between intentional and incidental learners lies in the number and distribution of associations of different strengths formed during training. Both groups form some strong associations, but incidental learners form fewer associations of intermediate strength and more weak associations than do intentional learners. Hence neither a very easy test (recognition or facilitating context at recall) nor a very difficult test (inhibitory context at recall) is sufficiently sensitive to discriminate between intentional and incidental learners. A test sensitive to intermediate degrees of associative strength (free recall) allows the differences in degree of learning to produce significant differences in performance.

References

1. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.
2. Haagen, C. H. Synonymity, vividness, familiarity and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1948, **27**, 453–463.
3. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
4. McGeoch, J. A., & McGeoch, G. O. Studies in retroactive inhibition: VI. The influence of relative serial position of interpolated synonyms. *J. exp. Psychol.*, 1937, **21**, 320–329.
5. Mood, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.
6. Postman, L., & Phillips, L. W. Studies in incidental learning: I. The effects of crowding and isolation. *J. exp. Psychol.*, 1954, **48**, 48–56.

(Received April 14, 1954)

Journal of Experimental Psychology
Vol. 49, No. 1, 1955

TRANSPOSITION IN THE FEEBLEMINDED

HAROLD W. STEVENSON AND IRA ISCOE

The University of Texas

Two studies have recently been published in which tests of transposition with "preverbal" children yield data which have been interpreted as confirming the theory of transposition proposed by Spence (**8**). Kuenne (**6**) and Alberts and Ehrenfreund (**1**) found that for young children the incidence of transposition decreased as the separation between the training and test stimuli increased. These results are predicted from Spence's theory and are in line with those obtained from studies with lower animals (**2, 5**). The responses of the older children were, in contrast, found to be "keyed to the words relating to the cue aspect of the stimuli" (**6**, p. 488) and transposition for near as well as far stimuli was nearly perfect.

The present experiment was designed to test the generality of these results with other "preverbal" *S*s, in this case feebleminded *S*s. The lack of verbal and conceptual control of responses is a characteristic frequently noted in the feebleminded, and such individuals should make excellent *S*s to provide further tests of the hypotheses of Kuenne that (*a*) the mechanisms assumed by Spence to underlie transposition behavior in animal *S*s are also operative in preverbal human *S*s, and that (*b*) transposition for these *S*s should decrease as the test stimuli become increasingly remote from the training pair.

METHOD

Subjects.—The *S*s were 51 feebleminded individuals residing at the Austin State School.[1] Their ages ranged from 9 to 40 yr. Only individuals having an IQ above 40 or an MA above 6 yr. were used, since it was found on the basis of pretests and previous experiments that *S*s below these levels took an inordinate amount of time to solve the problem. The individuals were all of the familial type of mental deficiency except for a few *S*s classified as feebleminded with undetermined etiology. No individuals with gross motor or sensory disturbances were used. Of the 51 *S*s, 44 completed the experiment. Two *S*s could not be obtained to complete testing, and five *S*s were dropped because of position or alternation habits which could not be broken after an average of 150 trials.

Materials.—The equipment consisted of a screen and a board with six blocks. The screen was used to prevent *S* from observing *E* arrange the blocks. The dimensions of the board were 14 by 9 in. Two holes 1¼ in. in diameter and separated by 3½ in. were drilled along the midline of one board and served to contain the reward. On three sides of each hole was a raised wooden guide to facilitate the proper placement of the blocks over each hole. Each block was 3 in. square and ½ in. high. The board and blocks were painted flat black. Mounted in the center of each block was a white cardboard square, the stimulus to be discriminated. The top of each block was covered with a piece of clear Lucite to prevent soiling through handling. The areas of the stimuli on Blocks 1 to 6 were ¼, ½, ¼, 1, 2, and 4 sq. in. respectively. The size ratio of each stimulus to the preceding one was 2 to 1. Small cinnamon "red hots" were used as a reward.

Procedure.—Each *S* was brought to the experimental room by a ward attendant. The *S* was told that this was a game and not a test, and was seated across from *E* as *E* demonstrated how the blocks could be placed over the holes and how they could be removed. He then told *S*: "Now see these blocks. I'm going to hide a piece of candy under one of these blocks and I want to see if you can find it. You pick the block you think the candy is under." No mention was made of the stimulus squares. The board was removed behind the screen and the blocks and candy arranged. The *E* placed the

[1] We gratefully acknowledge the cooperation of the staff of the Austin State School and especially appreciate the aid of the superintendent of the school, Dr. Charles W. Castner, and of Mr. Charles Clelland, the school's psychologist.

board in front of *S* saying, "Where's the candy?" The *S* was allowed to pick up only one block on each trial. The *E* attempted to make the game as enjoyable as possible and praised *S* when a correct choice was made. Motivation throughout the experiment was therefore kept quite high.

All *Ss* were trained with the blocks containing the two largest stimuli, Blocks 5 and 6. For all *Ss* the candy was placed under the block with the smaller stimulus square and the position of the stimuli on the board followed a prearranged random sequence of 20 positions. The criterion for learning was nine correct choices out of ten. Approximately 60 trials were given each day until the criterion was met. Immediately after *S* attained this criterion, the test trials were begun without comment.

For the test trials three groups were formed, and *Ss* were assigned to each at random. Group 1 was tested on Blocks 4 and 5; Group 2 on Blocks 3 and 4; and Group 3 on Blocks 1 and 2. The candy was again to be found under the smaller block of each pair. The test trials were continued until five consecutive correct responses were made or until 40 trials had been given. The possibility of fatigue and the limited availability of the *Ss* made this desirable. At the end of the experiment *Ss* were questioned concerning their basis of response.

Results

Training trials.—The mean number of trials excluding the ten criterion trials was 41.9, with an *SD* of 49.8. The minimum number of trials was zero and the maximum was 212. Both the large number of trials required on the average to solve this simple size discrimination problem and the great differences in speed of acquisition among the *Ss* are noteworthy. The poor performance of these *Ss* is evidenced by the fact that a group of 20 normal children with an average MA and CA of 7.8 yr. learned the identical discrimination with a mean of 28.6 trials and an *SD* of 25.6 (**7**).

The three experimental groups were quite comparable in terms of CA, MA, and trials required to achieve the initial discrimination, as may be seen in Table 1. When tested with Festinger's *F* test (**3**), none of the differences between the mean trials was found to be significant at even the 5% level. No significant relationship between either MA or CA and number of trials required for solution was found. The levels of significance resulting from appropriate χ^2 tests of such relationships were .20 or less.

Test trials.—Two commonly accepted criteria for assessing transposition are the initial test response and the responses made on the first several test trials. Using the first criterion for our combined groups, it is found that 29, or 66%, of the *Ss* chose the smaller stimulus and thereby may be judged to have transposed. This is significantly greater than the 50% that would be expected by chance ($CR = 2.13$). The probability of getting this difference is significant at greater than the .05 level of confidence (two-tailed test). If transposition is tested according to the mean number of relational responses made on the first five test trials, it is found that the mean number of smaller choices was 3.48, which is greater than the mean of 2.5 which would occur if the choices were

TABLE 1

Performance of *Ss* on Training Trials

Group	*N*	CA in Years		MA in Years		Trials to Criterion		
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Median
1	15	18.1	5.7	7.6	2.7	35.8	38.3	22.0
2	15	20.6	6.0	7.0	1.6	53.7	61.4	36.0
3	14	23.9	9.2	7.8	1.7	35.9	44.2	21.5

made solely on a chance basis. The level of significance of this difference cannot be meaningfully assessed, however, because of the differential reinforcement given on the test trials.

The critical question is whether differences in transposition resulted, depending upon the degree of remoteness of the test stimuli. As may be seen in Fig. 1, the percentage of correct choices on the first test trial is approximately the same for all groups. The difference of 3% between Group 3 and the other two groups is obviously not significant. Similarly, the percentage of "smaller" responses on the first five test trials is approximately the same for the three groups. The greatest difference, that between Groups 2 and 3, yields a *CR* of only .59. The results when judged by both these criteria, therefore, are in contrast to those of previous studies, for no decrease in transposition is found with increased remoteness of the test stimuli.

In one respect, however, the present data are coordinate with the prediction that transfer is inversely related to stimulus separation, and this is with regard to the ease with which *Ss* learned the correct discrimination on the test trials. In Group 1 the median number of trials required was 3.0, whereas it was 7.0 in Group 2 and 15.0 in Group 3. When the groups' relearning scores are compared by Festinger's *d* test (4), a significant difference is found between Groups 1 and 3 ($d = 3.6$; $p < .05$), but not Groups 2 and 3 ($d = 2.1$; $p > .05$) or Groups 1 and 2 ($d = .7$). Further, differences are seen in the number of *Ss* who showed perfect transposition by choosing the smaller stimulus on the first test trial and continuing to do so until the criterion of five correct responses was reached. Six *Ss* in Group 1, 5 in Group 2, and 3 in Group 3 performed in this fashion.

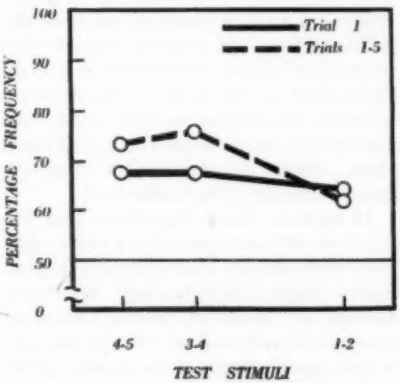


FIG. 1. The proportion of "smaller" choices on the first test trial and on the first five test trials, according to test stimuli.

Verbalization.—Careful attention was paid to the questioning of *Ss* concerning how they solved the problem. Only 3 of the 44 *Ss* were able to verbalize the principle involved. It is of interest to note that all three transposed in the first test trial. Two more *Ss* mentioned something about the squares but could go no further in their explanation. The remainder could say nothing concerning their solution, other than insisting they did not know or just guessed. Thus the general assumption that responses were not mediated verbally in these *Ss* seems a valid one. Even though their average MA's are higher than those of the verbal groups in the previous studies, the limited nature of their verbal and conceptual abilities is apparent.

DISCUSSION

It is difficult to conclude whether or not Kuenne's hypotheses concerning mode of learning or incidence of transposition in the preverbal *S* have been verified since differing results are obtained, depending upon the criterion of transposition adopted. For example, neither we nor Alberts and Ehrenfreund (1) obtained a gradually decreasing

gradient for transposition on the first test trial. For the first five test trials, however, others have found a gradient, while we did not. We did find a gradient of the predicted type for learning scores on the test trials and for the consistency of correct test response. The criterion used, then, may be a significant variable in determining whether or not a gradient is found.

There is evidence also that the gradient may be affected by several variables concerning the training and testing procedure. At least two differences between this and previous experiments may involve such variables.

The first difference is that the stimuli used were within a smaller range of absolute values, compared to those of previous studies. The smaller range may affect transposition by (*a*) reducing the within-pair size differences and thus increasing the difficulty of discriminating between the stimuli within each pair, or (*b*) resulting in stimulus generalization from the training pair such that the responses to the stimuli in each test pair have relatively the same habit strength. Concerning the latter, it may be easily shown with Spence-type curves that when the range of stimuli in absolute area is restricted, the net differences in effective excitatory strength of responses associated with successive stimuli are relatively small. As a consequence, no response to either stimulus of each stimulus pair has a clear dominance in habit strength and transposition would be slight.

For the present experiment, neither explanation seems to give an adequate account of the results. The first seems unlikely, since *S*s did equally as well on the early test trials on the near stimuli which had a relatively great size difference as on the far stimuli where the size difference was relatively small. The second is difficult to accept in view of the ease with which the original learning and especially the relearning were accomplished. According to this explanation, learning should have been much slower. The possibility remains, however, that transposition may be affected by the variable of stimulus range.

A second difference between the experiments involves the procedure used in testing transposition. In both of the previous studies, a 24-hr. delay was interpolated between the training and test series, while in the present study the test trials followed the training trials immediately. This variable might easily produce differences in transposition through its effects on the ability of *S* to discriminate between the training and test pairs of stimuli. A delay may increase this difficulty to such a degree that *S* is unable to make such a discrimination and continues therefore to make the same relative response to the test stimuli as he has made to the training pair. In contrast, immediate presentation of the stimuli would decrease the difficulty of discriminating between the stimulus pairs, and the probability that the discriminated stimuli would evoke the same relational response would be decreased. In brief, an extended time lapse between training and testing may reduce the discriminability of the training and test stimuli to such a degree that transposition is assured. Furthermore, a time lapse may affect transposition differentially, with its greatest effect being on the near stimuli. The far stimuli, especially when a size ratio of 2 to 1 is used, are so radically different from the training stimuli that even a day's pause would have relatively little effect on *S*s' ability to discriminate between the training and test pairs. An explanation in terms of the time factor, then, seems to provide a plausible account for the results obtained.

It appears likely, therefore, that the gradient of transposition is a more complex phenomenon than previous experiments might indicate, and that such factors as mode of testing, absolute range of stimuli used, and the presence of a time lapse between training and testing are variables that may be critical in determining its presence.

Summary

A group of 44 feebleminded *Ss* with an average MA of seven years was trained to choose the smaller of two stimulus squares. Transposition was tested with pairs of stimuli removed 1, 2, or 4 steps from the training pair. The results indicate:

1. Learning by the feebleminded *Ss* was slow, and only three of the *Ss* could verbalize their basis of response.

2. The over-all incidence of transposition was significantly above chance.

3. No decrease in transposition was found on the first test trials with increasing separation of training and test stimuli, but an inverse relationship was found for learning scores on the test trials and for consistency of correct test response.

The results are discussed in terms of the differences in the design of this and former experiments.

References

1. Alberts, E., & Ehrenfreund, D. Transposition in children as a function of age. *J. exp. Psychol.*, 1951, **41**, 30–38.
2. Ehrenfreund, D. A study of the transposition gradient. *J. exp. Psychol.*, 1952, **43**, 81–88.
3. Festinger, L. A statistical test for means of samples from skew populations. *Psychometrika*, 1943, **8**, 205–210.
4. Festinger, L. The significance of difference between means without reference to the frequency distribution function. *Psychometrika*, 1946, **11**, 97–105.
5. Kendler, T. S. An experimental investigation of transposition as a function of the difference between training and test stimuli. *J. exp. Psychol.*, 1950, **40**, 552–563.
6. Kuenne, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.
7. McConnell, C. A developmental study of transposition. Unpublished study, Univ. of Texas, 1954.
8. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

(Received April 21, 1954)

Journal of Experimental Psychology
Vol. 49, No. 1, 1955

PROBABILITY LEARNING IN A PROBLEM-SOLVING SITUATION[1]

JACQUELINE JARRETT GOODNOW AND LEO POSTMAN

Harvard University *University of California*

There is considerable experimental evidence that organisms can learn to discriminate probabilities. Such evidence has come from two classes of experiments. First, there are studies like those of Brunswik and Herma (**2**) and Levin (**11**) which demonstrate the acquisition of perceptual biases based on probabilistic cues. Perceptual learning occurs independently of *Ss*' ability to identify the cues which influence their judgments. A second class of experiments has investigated the choice behavior of *Ss* when confronted with alternatives each of which involves probability rather than certainty of outcome. Probability of reward as a determiner of choice behavior in rats was demonstrated by Brunswik (**1**). Human *Ss* have been asked to predict whether or not a light would flash on (**3, 4, 5, 7**), whether a vertical or horizontal design would be presented (**6**), or whether *E* would say "plus" or "check" (**9**). In these experiments each of the alternative events occurs with a probability greater than zero but less than 1. In general, close agreement between event probabilities and relative frequencies of choices has been found, although Brunswik and his associates have emphasized that such agreement should not be regarded as a necessary feature of probability learning (**2, 9**).

In experiments using simple prediction tasks, the *S* is apt to realize that he must make his responses on the basis of probabilities. In making his predictions, *S* may, therefore, be guided by his guesses and convictions concerning the nature of chance distributions, and particularly sequences and patterns within such distributions. Hake and Hyman suggest that *Ss* "do not perceive the probability rules by which a series of events was generated. They perceive, instead, those short sequences of events which precede each prediction" (**6**, p. 72). To the extent that the stimulus series provides discriminable sequences which serve as more or less valid cues to future occurrences, *S*'s predictions will conform to the actual probabilities of the events.

The question arises of whether probability learning will occur in choice situations in which *S* is not guided by his concepts of chance sequences but rather attempts to find a lawful solution to a problem. The present study attempts to provide an experimental answer to this question.

METHOD

Experimental task.—The *S*'s task was to match geometric designs. On successive trials, *S* was presented with a set of three cards on each of which appeared a pair of geometric designs. One of these cards was a "key card," and the other two showed variations of the designs on the key cards. On each trial, *S* was required to decide which of the two variations constituted a correct match to the key card.

The geometric designs were those used in a study of memory change by Postman and have been described elsewhere (**14**). There were ten sets of five cards each. Each set consists of a basic pair of designs and four variations. The variations are derived from the basic pair by

[1] This study was facilitated by a grant from the Laboratory of Social Relations, Harvard University. We wish to express our gratitude to Drs. J. S. Bruner, R. Bush, and F. Mosteller for generous advice.

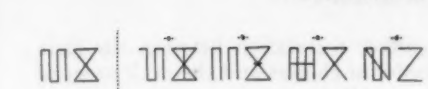


FIG. 1. Example of key card and variations. *a* and *b* are subtractive variations; *c* and *d* are additive variations.

removing a line from one member of the basic pair and adding it to the other. Thus all the elements of the basic pair are preserved in the variations. An example of a basic pair along with its four variations is shown in Fig. 1. As Fig. 1 shows, there are two variations (*a* and *b*) in which a line is removed from the first member of the pair and added to the second member, and two variations (*c* and *d*) in which a line is removed from the second member and added to the first. With reference to the first member of the pair, the variations follow the principles of subtraction and addition of lines respectively. These two types of variation provide the alternatives in the two-choice problem with which *S* was faced on each trial. Even though the specific designs kept changing from trial to trial, *S* always had to choose between an additive and subtractive variation.

The *S* was led to believe that the correct choice depended on the nature of the designs and that there was a general principle governing the correct choices. Since the designs varied widely in geometric properties, it was possible to form a wide variety of hypotheses linking additive and subtractive choices to characteristics of the designs. In fact, however, the correct choice was predetermined. The additive variation was called correct on a certain percentage of the trials and the subtractive variation on the remaining percentage of the trials.

Procedure.—The experiment was introduced as a study of learning. The *S* was shown a sample set of three cards, one key card and two variations, and the way in which variations can be identified as "adding" and "subtracting" was explained. A second sample set was then presented to check on *S*'s grasp of the distinction. After *S* had been instructed in the matching task, a practice series of ten unreinforced trials was presented. The responses during the practice series provided a rough estimate of *S*'s guessing preferences.

Following the practice series, there were 80 trials, divided into 8 blocks of 10 trials each, in which *S*'s choices were differentially reinforced according to a predetermined schedule. Each of the 10 key cards was presented once in a given block.

In presenting the stimuli, *E* drew the cards from a rack in a prearranged order. The key card was drawn and placed down first. Then the two variation cards were drawn and placed down simultaneously. After *S* had made his choice, *E* recorded the response and commented on it as right or wrong. The two variation cards were then withdrawn simultaneously and then the key card.

If *S*'s choice was correct on a given trial, *E* simply said "Right." If the choice was incorrect, *E* stated, "No, this 'adding' (or this 'subtracting') card is correct."

For each choice, *S* was limited to about 5 sec. after presentation of the cards. The interval between trials was the minimum time required for changing the set of stimuli. The interval between successive blocks was approximately 30 sec., except that after each second block *S* was asked: "How did you decide which card to choose?" The *S*'s responses were recorded in brief form by *E*.

Schedule of correct choices.—There were six experimental groups which differed in regard to the relative frequencies with which additive and subtractive variations were called correct, viz., 50:50, 60:40, 70:30, 80:20, 90:10, and 100:0.

Except for the 50:50 case, the additive variations always had the higher probability of being correct. The decision to favor the additive variations was based on the results of a pilot study which had revealed a general preference for subtractive variations at the beginning of the series. The *S*s who showed a preference for additive variations during the practice series were put into the 50:50 or the 100:0 group. The *S*s assigned to the other groups were matched for their initial guessing preferences.

The reasons for this deviation from a strict matching procedure were as follows: In order to test the learning of probabilities under the most stringent conditions possible, it was desirable to make the more profitable alternative opposite to that initially favored by *S*. This consideration does not, however, apply to the 50:50 condition. In the case of the 100:0 condition, preliminary experimentation had shown the learning task to be so easy that initial bias was not likely to have any effect. As it turned out, very few *S*s showed an initial preference for the additive alternative so that the groups varied only slightly in initial bias. Analysis of variance shows that the six experimental groups did not differ significantly in the number of guesses of the additive alternative during the practice series ($F = 0.22$). Any difference found among the conditions may, therefore, be ascribed to the experimental treatment.

The probabilities apply to each block of 10 trials. It must be stressed that the probabilities refer to the relative frequencies with which the two alternatives were scheduled to be correct,

regardless of *S*'s choice, rather than to proportions of reinforcement. How often *S* was reinforced depended on the frequency with which his choices coincided with the predetermined schedule of correct alternatives.

The schedule of correct alternatives involved restricted randomization since it was necessary to minimize the effect of memory for responses to specific designs. The same key card was not allowed to appear on trial 10 of one block and on trial 1 of the following block. Identical sets of three cards did not appear in two consecutive blocks. This safeguard was probably adequate since each set of three cards appeared only twice in a series of 80 trials. Within the limits of the probabilities scheduled for the various groups, additive variations were correct equally often for each key card over the whole series of trials.

Subjects.—The *S*s were undergraduate students at Harvard University. They did not know the purpose of the experiment. There were eight *S*s in each of the six experimental groups.

Results

Frequency of choices as a function of event probability.—Table 1 shows the average number of choices of the additive variation by blocks for all the experimental groups. At the end of the training period, there is in general close correspondence between the distribution of choices and the probabilities of alternative outcomes. This agreement is not perfect; for example, the 70:30 group splits its choices 60:40. There is, however, perfect correlation between the rank order of choices and the rank order of probabilities.

Inspection of the variations from block to block within experimental groups suggests that the speed with which the probabilities are learned varies from condition to condition. In general, the higher the proportion of the more profitable alternative, the faster is the rate of probability learning.

To test for the presence of significant differences among the experimental groups, a ranked one-way analysis of variance was performed on the number of additive choices made by *S*s over the last 40 trials. This method, developed by Kruskal and Wallis (**10**), makes no assumptions about the normality of the distribution of scores. The variation among groups was found to be highly significant ($H = 43.9$, $df = 5$, $p < .01$). Since the inclusion of the 100:0 group tends to exaggerate the difference among conditions, the analysis was repeated with the 100:0 group omitted. Variation among the remaining five groups was still highly significant ($H = 27.5$, $df = 4$, $p < .01$).

To evaluate the significance of dif-

TABLE 1

Number of Choices of the Additive Variation by Blocks of Trials for the Different Experimental Groups

Block	Ratio of Additive to Subtractive Reinforcements											
	50:50		60:40		70:30		80:20		90:10		100:0	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	3.75	1.71	5.00	1.41	4.25	1.92	4.88	2.19	4.75	2.54	5.62	1.43
2	4.12	1.71	4.38	1.79	4.75	1.85	4.38	1.71	8.12	1.99	8.50	1.58
3	5.50	1.58	5.38	0.66	5.62	1.34	6.00	1.80	9.12	1.40	9.00	1.32
4	4.38	1.30	5.00	1.87	6.12	1.98	7.00	1.87	9.25	0.66	9.88	0.10
5	5.12	2.17	4.88	1.95	6.00	1.50	7.38	1.91	9.25	0.83	10.00	0.00
6	5.12	1.71	5.25	1.39	7.50	1.50	8.12	1.20	9.12	0.97	10.00	0.00
7	4.25	1.20	5.25	1.39	6.38	1.30	7.75	1.30	9.25	0.66	10.00	0.00
8	5.25	0.83	5.75	1.92	6.00	0.87	8.25	1.64	9.38	0.63	10.00	0.00

ferences between individual conditions, the method of allowances developed by Link and Wallace (cf. **13**) was used. This method uses ranges as a measure of variation and makes it possible to set allowances—essentially confidence limits—on differences between a number of groups. Since the method assumes homogeneity of variance, the 100:0 group, which has no variance, was omitted. Application of this method reveals significant differences at the .05 level of confidence between all groups except for the difference between the 50:50 and 60:40 groups. The difference between the 100:0 and 90:10 groups was significant at the .05 level of confidence by the Mann-Whitney test (**12**).[2] The statistical analysis confirms the close relation between choice distributions and event probabilities.

Relation of choices to verbalized hypotheses.—Although *Ss*' responses conformed to the probabilities of alternative events, the reasons given for the choices indicated that *Ss* did not recognize that the problem was one of probability discrimination. Furthermore, the reasons stated by *Ss* bore little relation to the actual distributions of choices. A wide variety of different hypotheses was offered by *Ss* showing the same distribution of choices. A given *S*, moreover, would frequently change his verbalized hypotheses while maintaining a fairly consistent distribution of choices from block to block. It should not be concluded that *Ss*' responses were entirely independent of the verbalized reasons for making the choices. Rather, the stated hypotheses had only "local" effects, i.e., they influenced some particular choices or short sequences of choices but did not determine the general form of the choice distribution.

There is additional evidence in support of the conclusion that the learning of probabilities developed "without awareness." Only three of the 40 *Ss* in the partial reinforcement groups abandoned the search for a lawful solution based on the nature of the designs and treated the task as one of discriminating probabilities. Finally, analysis of *Ss*' response sequences shows no evidence of the type of strategy that may be expected of *Ss* with knowledge of the probability pattern.

There is one type of error which *Ss* with knowledge of the probability pattern can easily avoid, viz., that of persisting in a given kind of response after the expected number of times it can be correct in a given block has been exhausted. If *S* realizes, for example, that "adding" is correct eight times and "subtracting" two times in each block, then he should not choose "subtracting" after the latter has been correct twice. The frequency of such "avoidable errors" was determined for the 80:20 and 90:10 groups, i.e., for the cases in which recognition of the probability pattern would be easiest. As Table 2 shows, there is no systematic decrease in the frequency per trial with which "subtracting" choices were given after

TABLE 2

FREQUENCY PER TRIAL WITH WHICH SUBTRACTIVE VARIATION WAS CHOSEN BEFORE AND AFTER POSSIBLE NUMBER OF CORRECT CHOICES HAD BEEN REACHED
(Data for Blocks 5–8)

Condition	Before	After
80:20	1.67	1.78
90:10	0.50	1.00

[2] As Brunswik has suggested, the difference between the 90:10 and the 100:0 conditions may reflect *Ss*' disposition to draw a sharp line between exceptionless rules and statistical regularities of high probability (**1**, p. 188 f.). Brunswik did not find such a tendency in the case of animal *Ss*.

the point of avoidable error was reached. One final piece of evidence for *Ss*' failure to perceive the task as a probability problem is the absence of sequences or "runs" of responses such as did occur in a situation which was objectively identical but was presented as a gambling task (**8**).

The role of memory.—Each of the ten key cards appeared once in each block. The *Ss* might well remember from one block to the next which response had last been correct for a particular key card and repeat the previously correct response when that key card appeared again.[3] Such a procedure would lead to many errors since the schedule of correct choices was so set up that the same answers were seldom correct for the same key cards in two consecutive blocks. A deliberate strategy of making choices on the basis of memory for past successes and failures would soon have to be abandoned. It is possible, nevertheless, that such a strategy was used by some *Ss* for at least part of the time. Moreover, we cannot disregard the possible "automatic" effects of reinforcements in establishing associations between particular key cards and particular responses.

Choices based on memory for past responses to particular designs would produce the following sequences from one block to the next: (*a*) repetition of a rewarded response to a particular key card; (*b*) avoidance of a punished response to a particular key card. To determine whether such sequences occurred at a better than chance level, the procedure used in evaluating the frequency of correct hits in a matching problem was applied.[4] We define as a "hit" each trial on which *S*'s choice involved one of the two sequences described above. The expected number of hits $E(h)$ and the expected standard deviation of this number σ_h was determined for each pair of sequential blocks for each *S*. The significance of the difference between expected and obtained hits is evaluated by means of a *t* test. When the values of *t* are summated for all blocks and all *Ss* in a given condition, the results shown in Table 3 are obtained.

TABLE 3

SIGNIFICANCE OF DEVIATION FROM CHANCE OF FREQUENCIES OF CHOICES POSSIBLY BASED ON MEMORY FOR PAST RESPONSES

Condition	Value of Summated *t*	*p*
50:50	2.39	.02
60:40	0.77	—
70:30	1.32	—
80:20	2.82	.01
90:10	2.34	.05

In three of the five conditions, responses which *may* be ascribed to memory for the last response to a particular key card occur at a better than chance frequency. In many cases, of course, conditions other than memory may have produced the "hits." The fact that in two of the conditions, in which probability learning did occur, the number of "hits" is not significant, strengthens the conclusion that memory (or automatic reinforcement effects) was at best only one of several factors influencing *Ss*' choices.

DISCUSSION

The results of the experiment present evidence for probability learning in a situation in which *Ss* do not attempt to discriminate probabilities but rather search for a lawful solution to a problem. Our findings serve to give greater generality to the fact of probability learning.

[3] Memory for specific sets of three cards can be discounted since such identical sets appeared only twice in the total series.

[4] We are grateful to Dr. F. Mosteller for suggesting this analysis.

It can be inferred from the distributions of choices that *Ss* discriminated the probabilities of alternative events; yet their verbal responses gave no indication that they perceived the problem as one of probability discrimination. In agreement with previous studies of probability learning (**2, 9**) our findings constitute positive evidence for "learning without awareness." At the same time, it should be recalled that there are other situations in which awareness of the principle governing correct responses appears to be an essential condition of improvement (**15**). The question as to whether improvement in complex cognitive tasks depends on awareness of the principle of solution does not have an all-or-none answer.

Little is known concerning the specific cues mediating the discrimination of probabilities. In the case of simple prediction tasks, in which the nature of the stimuli does not vary, *S* may be responding to "those short runs of events which precede each prediction" (**6**, p. 72). In our situation, short runs of events are probably less effective as cues to responding. Since the stimulus materials change in a number of respects from one trial to the next, we should expect cues of temporal order to be less effective than they are in a situation in which stimuli differ only in temporal order. Thus cues derived from the geometric properties of the designs would interfere with the effectiveness of cues based on temporal order. Examination of our *Ss*' response sequences confirms this conclusion. In previous studies using invariant stimuli the cue value of short runs was demonstrated by "recency effects," i.e., systematic increases or decreases in the use of a given alternative following a short series of successful predictions (**2, 6, 9**). There was no evidence for such recency effects in the response sequences of our *Ss*. It appears that *Ss* gradually came to respond to the total probability pattern. Our data do not allow us to speculate further about the ways in which a probability pattern becomes translated into effective cues for sequences of choices. As in other cases of psychophysical discrimination, these is no reason to expect a one-to-one correspondence between stimulus difference and relative frequency of choice. Such one-to-one correspondence is, indeed, not present in our data even though the rank-order correlation between probabilities and frequencies of choice is perfect.

Summary

This experiment demonstrates the occurrence of probability learning in a problem-solving situation. Confronted with a two-choice task, *Ss* learned to respond in accordance with the probabilities of alternative outcomes even though they did not recognize the task as a probability situation and attempted to find a lawful solution to the problem.

The experimental task consisted of matching geometric designs. On each trial *S* had to choose between two types of variation as matching a key design. There were six experimental groups which differed in the relative probabilities of the two types of variation being correct, viz., 50:50, 60:40, 70:30, 80:20, 90:10, 100:0. There was a high correlation between the probabilities of alternative outcomes and the relative frequencies of choices. In the large majority of cases *Ss* learned to discriminate the probabilities without becoming aware that the problem was one of probability discrimination.

References

1. Brunswik, E. Probability as a determiner of rat behavior. *J. exp. Psychol.*, 1939, **25**, 175–197.
2. Brunswik, E., & Herma, H. Probability learning of perceptual cues in the establishment of a weight illusion. *J. exp. Psychol.*, 1951, **41**, 281–290.
3. Calvin, J. S. Learning of probable occurrences. *Amer. Psychologist*, 1948, **3**, 236. (Abstract).
4. Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.
5. Grant, D. A., Hornseth, J. P., & Hake, H. W. The influence of the inter-trial interval on the Humphreys' random reinforcement effect during the extinction of a verbal response. *J. exp. Psychol.*, 1950, **40**, 609–612.

6. HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.
7. HUMPHREYS, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
8. JARRETT, J. Strategies in risk-taking situations. Unpublished Ph.D. thesis, Radcliffe College, 1951.
9. JARVIK, M. E. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, 1951, **41**, 291–297.
10. KRUSKAL, W. H., & WALLIS, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. statist. Ass.*, 1952, **47**, 583–621.
11. LEVIN, M. M. Inconsistent cues in the establishment of perceptual illusions. *Amer. J. Psychol.*, 1952, **65**, 517–532.
12. MANN, H. B., & WHITNEY} D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
13. MOSTELLER, C. F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley Press, 1954.
14. POSTMAN, L. Learned principles of organization in memory. *Psychol. Monogr.*, 1954, **68**, No. 3 (Whole No. 374).
15. POSTMAN, L., & JARRETT, R. F. An experimental analysis of 'learning without awareness.' *Amer. J. Psychol.*, 1952, **65**, 244–255.

(Received April 12, 1954)

Journal of Experimental Psychology
Vol. 49, No. 1, 1955

THE RELATION OF CONDITIONED DISCRIMINATION TO THE MMPI *Pd* PERSONALITY VARIABLE[1]

ANNE BONNER WARREN AND DAVID A. GRANT

University of Wisconsin

This experiment was designed to investigate the relation between conditioned discrimination and the personality characteristic of psychopathic deviation as defined by the Minnesota Multiphasic Personality Inventory (MMPI) (**8**).

Individual and personality differences in the ease of conditioning and in other conditioning and learning phenomena have interested workers in this area since the time of Pavlov (**1, 2, 3, 4, 5, 12, 14, 16, 17, 19**). Some of this interest has centered about conditioned discrimination (**11, 15**). Present-day investigators are no longer content with the broad statement that individual differences exist or with the typological approach proposed by Pavlov (**12,** pp. 304–308); instead they tend toward more analytic approaches. These approaches are represented by Hilgard, Jones, and Kaplan (**11**), who hypothesized that *Ss* showing greater anxiety would fail to eliminate responses to the negative stimulus in conditioned discrimination, and by Spence and Farber (**15**), who hypothesized that anxiety should lead to higher excitatory potential and better discrimination between positive and negative stimuli. Also relevant is the study by Bitterman and Holtzman (**2**), who predicted and observed higher average response level for "anxious" *Ss* during acquisition and extinction of the conditioned GSR.

[1] This investigation was supported in part by the Research Committee of the Graduate School of the University of Wisconsin with funds supplied by the Wisconsin Alumni Research Foundation.

The present study is concerned with another facet of the problem. It investigates the hypothesis that *Ss* who score high on the psychopathic deviation (*Pd*) scale of the MMPI will show atypical conditioned discrimination. It was predicted that even normal college students, such as the *Ss* of this study, who score high on the *Pd* scale will tend to seek immediate gratification and avoid immediate annoyances, social pressure notwithstanding. It was predicted that these *Ss*, having learned an avoidance response to an annoying stimulus (the corneal air puff), would fail to develop good conditioned discrimination because they would continue to respond to the negative stimulus, thus avoiding possible discomfort. This should occur in spite of instructions which lead to good discrimination in *Ss* with lower *Pd* scores.

Method

Apparatus.—Except for the presence of two milk glass windows for the presentation of the conditioned stimuli (CS) the apparatus was precisely the same as that used in previous studies from the Wisconsin laboratory (**6, 7**). The CS were similar to, though not identical with, those of the Stanford studies (**9, 10**). A small milk glass disk served as the fixation spot, and the two CS windows were on either side of this spot. They were called the left CS and the right CS, respectively. The CS consisted of illuminating one of these windows to a brightness of 215 to 230 mL., for a duration of about 750 msec. The UCS was a corneal air puff which was strong enough to evoke a sharp reflex blink. The *Ss* were run under standard light-adapted conditions with standard time intervals for conditioning (**6, 7**).

Procedure.—On Day 1 all *Ss* were given 60 reinforced trials, using one of the conditioned

stimuli and the UCS. On Day 2, 30 more reinforced trials were given using the same CS as on Day 1, but interspersed with these reinforced trials were 30 unreinforced trials with the other CS. The reinforced stimulus was called the positive stimulus, and the unreinforced stimulus was called the negative stimulus, using standard methods of contrasts training (12).

Experimental design.—For half of the *Ss* the right CS was positive and for the other half the left CS was positive. In addition each of these groups was subdivided so that half of the *Ss* had high *Pd* scores on the MMPI and half of them had low *Pd* scores. The design was thus a 2×2 factorial design with the right or left CS as one factor and high or low *Pd* scores as the other factor. The high-*Pd* *Ss* all scored 20 or higher on the revised *Pd* scale (65th percentile or higher) and the low-*Pd* *Ss* scored below the 65th percentile on the same scale. The *Ss* were run, and the records were read without *E*'s knowing the *Pd* scores.[2]

Subjects.—The *Ss* were men and women students in elementary courses in psychology who had taken the MMPI when they entered the University as freshmen. The *Ss* were selected randomly from within the MMPI groups and were contacted by telephone and asked to serve in the experiment. All but one *S* thus contacted cooperated in making appointments, but four *Ss* failed to appear for one or both sessions. A total of 23 men and 25 women composed the sample, and these were approximately evenly distributed with respect to the two independent variables in the experimental design. All *Ss* were told to look at the center fixation point, to remain as quiet as possible, and *not to control voluntarily their natural responses to the stimuli.*

Results

A conditioned response (CR) was defined as any eyelid response occurring between .25 and .50 sec. after the onset of the CS. In Fig. 1 is shown the percentage frequency of CR's plotted against successive blocks of ten trials for the two high-*Pd* groups and the two low-*Pd* groups separately. In Fig. 1 the high-*Pd* and the low-*Pd* groups are essentially alike with respect to acquisition on Day 1. This parallels the results of Hilgard, Jones, and Kaplan (11) on manifest anxiety, but does not agree with other anxiety studies (**14, 15, 16, 17**). On Day 2, however, the high-*Pd* group tended to give a higher frequency of responses than the low-*Pd* group, but more importantly, the high-*Pd* group tended to give a strikingly higher frequency of CR's to the negative stimulus than was the case with the low-*Pd* group. In other words, the striking feature of Fig. 1 is the fact that the difference in frequency of responses to the positive and to the negative stimuli is greater in the low-*Pd* group than in the high-

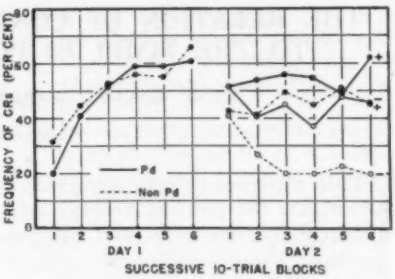


Fig. 1. Acquisition of eyelid CR (Day 1) and conditioned discrimination (Day 2) in high-*Pd* and low-*Pd* groups. Solid lines are for high-*Pd* groups; dashed lines are for low-*Pd* groups. Closed circles are for CR's to the positive (reinforced) stimulus and open circles are for CR's to the negative (unreinforced) stimulus.

TABLE 1

Summaries of Analyses of Variance of Frequency of CR's to the Positive Stimulus and the Negative Stimulus

Source of Variation	*df*	*F*'s on Positive Stimulus	*F*'s on Negative Stimulus
Pd score	1	1.35	6.02*
Position of positive stimulus	1	0.60	4.71*
Interaction	1	0.26	0.73
Error mean square	44	(851)	(649)

*Significant at 5% confidence level.

[2] The writers are indebted to Noel F. Kaestner who ran ten of the *Ss*.

Pd group. Table 1 presents the summaries of two analyses of variance on the total frequency of CR's to the positive and to the negative stimuli on Day 2. In Table 1 it can be seen that the stimulus position and the *Pd* score did not affect significantly the frequency of CR's to the positive stimulus, but that both the stimulus position and the *Pd* variable produced significant *F*'s for the differences in responses to the negative stimulus. In other words the heightened frequency of CR's to the negative stimulus in the high-*Pd* group, as shown in Fig. 1, was statistically significant. This indicated that there was statistically better discrimination between the positive and negative stimuli in the low-*Pd* group than was the case in the high-*Pd* group. In addition there was a tendency for all *Ss* to give a higher frequency of response to the right light than to the left, particularly when the right light was the negative stimulus. This result was unexpected and will be discussed later.

In order to analyze the discrimination process further a *relative discrimination score* was set up as the ratio of the number of CR's evoked by the positive stimulus to the total number of CR's during Day 2. This score was computed for each *S*, separately, in terms of all 60 trials on Day 2 and also separately for the three successive blocks of 20 trials. A high score indicated good discrimination. This discrimination score turned out to be highly variable from *S* to *S*. An analysis of variance was carried out on the relative discrimination scores for the 60 trials of Day 2 and for the three 20-trial blocks. The results of the over-all analysis are summarized in Table 2 where it is seen that the only significant difference was produced by the stimulus position, which gave an *F* significant at the 5% confidence level. No significant *F*'s were obtained for the nine comparisons in the three 20-trial blocks. The results of the analysis on *relative* discrimination scores did not, therefore, parallel the analysis of *absolute* response frequency summarized in Table 1. In part this may have been due to the notorious variability of ratios.

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE OF RELATIVE DISCRIMINATION SCORES

Source of Variation	*df*	*F*'s for Discrimination on Trials 1–60
Pd score	1	0.02
Position of positive stimulus	1	6.91*
Interaction	1	0.03
Error mean square	44	(405)

* Significant at 5% confidence level.

DISCUSSION

The results were in line with the hypothesis that the experiment was designed to investigate, namely, that there is a relationship between psychopathic deviation as measured by the MMPI and failure to develop good conditioned discrimination. This finding was based solely upon the responses to the negative stimulus to which the high-*Pd* group gave more than did the low-*Pd* *Ss*. No significant variation due to the personality variable was found in the responses to the positive stimulus. These results formed the same pattern as those of Hilgard, Jones, and Kaplan (11) in relation to anxiety, but they are rather unlike the results of Spence and Farber (15) on the anxiety variable.

The parallel between the results of the present study and the findings of Hilgard, Jones, and Kaplan made it desirable to test the extent to which the Taylor anxiety score and the MMPI *Pd* score measure the same thing. The correlation between the measures was found to be +.51 for 201 students in elementary

psychology courses at the University of Wisconsin. Thus, 25% of the *Pd* variance could be accounted for by the anxiety variable. This seems hardly enough to account for the common findings. It must be pointed out, of course, that both *Pd* and anxiety variables were sampled rather than manipulated in the conditioning studies, and it is hazardous to conclude that causal relationships exist between either anxiety or psychopathic deviation, per se, and discrimination. Both variables are highly correlated with other personality measures (**3, 8, 13**), and some of the related personality variables may be more basic to conditioned discrimination. It must suffice to say that the major findings of this study were in accord with the a priori prediction that high-*Pd* *Ss* would be unwilling to expose themselves to possible slight discomfort from the UCS despite social pressure from *E*'s instructions. Finally, it may be asserted that the basic assumption of this experiment suggests a rationale for the similarities observed between *Pd* and anxiety variables. That is, it may be that the seeking of immediate gratification and avoidance of immediate discomfort is the critical factor with respect to both variables.

The unexpected finding that better conditioned discrimination was obtained with the right light as the positive stimulus than with the left light deserves brief mention. This was especially true for the high-*Pd* group although the *Pd* × Stimulus Position interaction was not statistically significant. Three factors may have accounted for this difference: (*a*) The puff was delivered to the right cornea, and it was the right eyelid to which the thread was attached for recording the eyelid responses. (*b*) The controlling and recording apparatus was located to the right of *S*, and although *Ss* were never explicitly apprised of this, it is likely that they were aware of the presence and approximate position of the apparatus and of *E*. Any threat to *S* implicit in equipment or *E* would be localized to the right. (*c*) There was a very small difference in illumination between the right and left conditioned stimuli, the right stimulus being brighter by about 15 mL. It seems unlikely, however, that this difference should have contributed significantly to the frequency of CR's (**7**). On the whole, however, it seems unwise to emphasize unduly the results related to the position of the CS, because these results were unexpected and were significant at only the 5% level so that there is the distinct possibility of a Type I error.

Summary

The relationship between conditioned eyelid discrimination and the personality variable psychopathic deviation as defined by the MMPI was investigated in a sample of 48 men and women students from elementary psychology courses. The conditioned stimuli were lights to the left or right of the median plane and the USC was a corneal air puff. All *Ss* were given 60 standard conditioning trials on Day 1; on Day 2 they were given 30 trials with the positive CS of Day 1 and 30 unreinforced trials, randomly interspersed, using the other CS. Half of the *Ss* were conditioned to the right CS and half were conditioned to the left CS; half of each of these groups consisted of *Ss* scoring higher than 20 on the MMPI revised *Pd* scale and half who had scored lower than 20 on the MMPI revised *Pd* scale.

During Day 2 the average percentage frequency of CR's to the positive stimulus increased for all groups and the average percentage frequency of CR's to the negative stimulus decreased, an indication of conditioned discrimination. The discrimination was least in the high-*Pd* group where the frequency of the responses to the negative stimulus did not decrease as clearly as was true of the low-*Pd* *Ss*. This finding was significant at the 5% level, as had been predicted on the grounds that *Ss* who have high *Pd* scores would be more likely to preclude the possibility of immediate personal discomfort and would continue to blink to the negative stimulus.

References

1. Ausubel, D. P., Schiff, H. M., & Goldman, M. Qualitative characteristics in the learning process associated with anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 537–547.
2. Bitterman, M. E., & Holtzman, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 615–623.
3. Deese, J., Lazarus, R. S., & Keenan, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.

4. FARBER, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.

5. GAIER, E. L. Selected personality variables and the learning process. *Psychol. Monogr.*, 1952, **66**, No. 17 (Whole No. 349).

6. GRANT, D. A., & NORRIS, E. B. Eyelid conditioning as influenced by the presence of sensitized Beta-responses. *J. exp. Psychol.*, 1947, **37**, 423–438.

7. GRANT, D. A., & SCHNEIDER, D. E. Intensity of the conditioned stimulus and strength of conditioning: I. The conditioned eyelid response to light. *J. exp. Psychol.*, 1948, **38**, 690–696.

8. HATHAWAY, S. R., & MCKINLEY, J. C. *The Minnesota Multiphasic Personality Inventory*. (Rev. Ed.) New York: Psychological Corp., 1943.

9. HILGARD, E. R., CAMPBELL, A. A., & SEARS, W. N. Conditioned discrimination: the development of discrimination with and without verbal report. *Amer. J. Psychol.*, 1937, **49**, 564–580.

10. HILGARD, E. R., CAMPBELL, R. K., & SEARS, W. N. Conditioned discrimination: the effect of knowledge of stimulus-relationships. *Amer. J. Psychol.*, 1938, **51**, 498–506.

11. HILGARD, E. R., JONES, L. V., & KAPLAN, S. V. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94–99.

12. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning*. New York: D. Appleton-Century, 1940.

13. HOLTZMAN, W. H., CALVIN, A. D., & BITTERMAN, M. E. New evidence for the validity of Taylor's manifest anxiety scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 853–854.

14. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 116–119.

15. SPENCE, K. W., & FARBER, I. E. The relation of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, **47**, 127–134.

16. SPENCE, K. W., & TAYLOR, J. A. Anxiety and the strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

17. TAYLOR, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

18. TAYLOR, J. A. A personality scale for manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

19. TAYLOR, J. A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

(Received April 24, 1954)

Journal of Experimental Psychology
Vol. 49, No. 1, 1955

LEARNING SETS FROM MINIMUM STIMULI[1]

ARTHUR J. RIOPELLE

Emory University

The phenomenon of progressive improvement in intraproblem learning has important implications for all who attempt to maximize learning efficiency. It illustrates not only the extreme proficiency in generalized problem solution attainable by sophisticated primates, but also the slowness with which generalized learning efficiency develops. This phenomenon, called "deutero-learning" (**1**), the "formation of learning sets" (**2**), or "learning to learn" (**1, 2, 7**), has been demonstrated to occur in discrimination learning (**2, 7**) and will probably be found in social behavior (**1, 4**) and in the therapeutic situation (**5**). Since desirable behavior in these realms requires maximal adaptability of response to a changing environment, an analysis of the factors determining this flexibility is clearly needed.

Thus far, clearest demonstrations of the effect have involved discrimination learning in primates. Two diverse techniques have been employed. According to one technique, the rewarded member of a single pair of stimulus objects is successively alternated, usually after each problem is learned to a criterion (**7**). Under the second technique, new stimuli are introduced for each problem.

Since the latter technique, by its nature, demonstrates generalized efficiency, inquiry should be directed toward the significance and nature of the role played by the multiplicity of stimuli. Two contrasting possibilities suggest themselves: one emphasizing breadth of experience, per se, the other placing emphasis upon transfer relations among successive problems. The present experiment concerns the latter hypothesis. It aims to define the nature and to depict the course of interproblem transfer and to determine whether proficient performance on a series of problems generated from a minimum number of stimuli will also result in generalized discrimination learning and discrimination reversal learning.

[1] Supported in part by a research grant (M-589) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

METHOD

Subjects.—Five naive adolescent rhesus monkeys (No. 2, 6, 11, 15, 16) participated in this experiment. The *S*s were maintained and tested in the primate laboratory of the psychology department at Emory University. All were tamed and all were trained to displace a single stimulus object to secure food reward. All *S*s also learned ten preliminary discrimination problems prior to this experiment.

Apparatus and test-trial procedures.—The *S*s were tested in a modified version of the Wisconsin General Test Apparatus (**2**). In this apparatus the monkey confronts a sliding tray bearing two test objects. Each object covers a food well. If the correct object is displaced, the monkey secures a raisin or peanut reward; if the incorrect object is selected, the tray is retracted and reset for the next trial. This noncorrection procedure was used throughout the experiment. A blind and a one-way vision screen were appropriately employed to minimize artifacts.

Experimental design.—The experiment was conducted in two phases. During the first phase *S*s were tested on all possible problems that could be generated from only four stimulus objects. Since there are six combinations of four stimuli taken two at a time, and each combination yields two problems (by reversing polarity of reward), a total of 12 distinct problems was constructed. Table 1 shows a representative sequence of the 12 possible problems. This table also shows the relationships occurring between each problem and the one immediately

TABLE 1

ILLUSTRATIVE CATEGORIES OF PROBLEMS IN ONE SEQUENCE

Stimuli		Zero Displacement	Unit Displacement	
+	−			
A	B	Old +	All New	All New
A	D	Old −		
C	D	+ to −		Old +
B	C	− to +	+ to −	
C	A	All New		Reversal
D	B	All New	− to +	
A	C	+ to −		Old −
B	A	All New	Old +	
D	C	+ to −		− to +
B	D	+ to −	− to +	
C	B	All New		
D	A			

following it (zero displacement) and with the problem following the immediately successive one (unit displacement). Although 12 different sequences of problems were used, in none did a particular problem and its direct reversal succeed each other. In other words, at least one stimulus object was changed on every problem; furthermore, the introduction of a new stimulus provided no clue to the solution of the problem. The relations among successive problems were such as to provide two possibilities of partial reversal, − *to* + and + *to* −, and two possibilities of partial perseveration, *old* + and *old* −, as well as a complete change in stimuli, *all new*. Complete reversal is permitted only after an intermediate problem has intervened. This experiment thus contains nearly all possible transfer relations, yet holds the number of stimuli to an absolute minimum. Transfer relations, moreover, are maximized, since transfer based on identity of stimuli is substituted for transfer based on similarity.

All *Ss* were tested on 18 sequences of the 12 problems for a total of 216 problems. Each problem had to be learned to a criterion of five consecutive correct responses, but no problem was allowed to exceed 50 trials. Also, no new problem was begun if *S* had already completed 40 trials on that day.

During the second phase of the experiment *Ss* were tested for ten days on standard discrimination and discrimination reversal problems which involved stimuli never seen before. The procedures for this phase will be described in the following section.

RESULTS[2]

Four-stimulus discrimination learning.—Figure 1 shows the principal results of the first phase analyzed in terms of the possible transfer relations. Overnight transfer relations are omitted from the analysis. The points along the abscissas refer to consecutive blocks of 36 problems which are approximately uniformly distributed among the various categories of problems. Each curve of the left panel denotes a possible transfer relation between consecutive problems (zero displacement) whereas the curves of the right panel refer to transfer across an intervening problem (unit displacement).

Inspection of the left panel reveals that: (*a*) All curves descend toward the baseline. (*b*) The two partial reversal curves show negative transfer effects, whereas the partial perseveration curves show positive transfer effects. (*c*) The most efficient positive transfer situation occurs when the negative member of one problem is also negative on the next problem. (*d*) Differences among categories are greatest in the beginning of training but decrease with further training. These interpretations are supported by the analysis of variance, which showed highly significant (.001 level) differences among categories, blocks, and *Ss*. The only interaction term to reach significance was that for categories by blocks, which was significant at the .01 level.

Considering the right panel in rela-

[2] I am indebted to Mr. Elton L. Copelan for assistance in collecting these data.

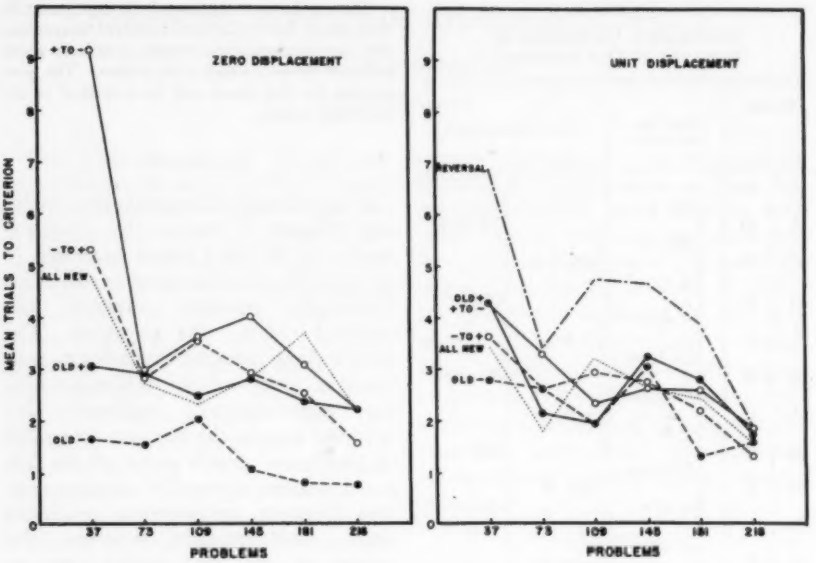


FIG. 1. Progressive discrimination performance analyzed in terms of transfer relations. The left panel refers to transfer from one problem to the next, the right panel refers to transfer across an intervening problem.

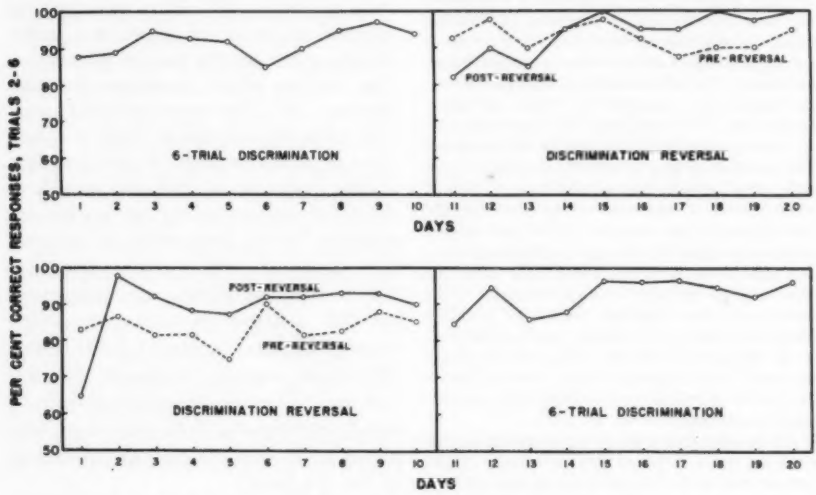


FIG. 2. Generalized discrimination and discrimination reversal performance after extensive training on four-stimulus learning.

tion to the left, two changes may be noted. First, the curves which were so different in the left panel follow essentially identical courses here. Second, the reversal curve is consistently the highest. Analysis of variance of these data showed only the main effects to be significant. In fact, under these conditions, 90% of the sum of squares due to categories of problems was contained in the single comparison (1 *df*) between the reversal curve and the mean of the five other curves. This comparison was highly significant (.001 level).

Generalized discrimination and discrimination reversal learning.—After completing the first phase of the experiment, *S*s No. 2 and 6 were trained on ten new six-trial discrimination problems each day for ten days. Following this they were trained on four discrimination reversal learning problems for ten days. Under this condition each problem is presented for six or eight trials in the manner of a standard discrimination problem. After the sixth or the eighth trial the reward is shifted to the other object without warning. Six postreversal trials are given. Subjects No. 11, 15, and 16 were tested first on discrimination reversal for ten days, then on standard six-trial discrimination learning for ten days.

The results are summarized in the graphs of Fig. 2. The panels of this figure show the performances on six-trial discrimination and discrimination reversal learning in the order in which they were administered. Each point represents the percentage of correct responses made during Trials 2–6 for the day indicated on the abscissa. For discrimination reversal problems both pre- and postreversal performances are shown.

A number of observations can be made from these graphs: (*a*) Discrimination learning performance is highly efficient from the first day forward. (*b*) Postreversal performance is superior to prereversal performance on Days 2–10. (*c*) Discrimination learning and discrimination reversal learning maintain this high level of efficiency during the second ten days of the testing.

Discussion

Four-stimulus learning.—The present experiment has provided an analysis of a succession of interproblem transfer relations. In general, the results have demonstrated the conditions which yield positive and negative transfer effects, the initial strength and subsequent decline of transfer through successive problems, and the limitations of its persistence across intervening problems. Positive transfer from one problem to the following occurs when the role of a stimulus object remains unchanged in both problems. Negative transfer, in contrast, occurs when its role changes in polarity from one problem to the next. Perfect symmetry is not to be expected, however; indeed, differential transfer effects can be found within the conditions making for positive and negative transfer.

A number of comparisons can be made between the present analysis and Spence's analysis of a related experiment (9). In that experiment he trained 12 chimpanzees on two separate difficult discriminations and then tested them on various pairings of the stimuli comprising the two preliminary problems. Spence's analysis involved the total number of previous reinforcements (those earlier than the preceding problem were reduced by only 10%) whereas this analysis involved only the previous two problems. It might be noted in this connection that only direct reversal effects were evident after unit displacement. Also, Spence's analysis considered only unreinforced wrong choices (frustrations) whereas the present analysis did not differentiate between unreinforced stimulus exposures and unreinforced selections of wrong stimuli. This latter difference is especially important since the most efficient transfer situ-

ation was that which depended upon the perseveration of the negative stimulus. Under these conditions, frustrations due to unrewarded responses occur only rarely, sometimes not at all for two or more successive problems, particularly during the later stages of training. Exposures to negative stimuli necessarily, however, equal those to positive stimuli.

Generalized discrimination and discrimination reversal.—The present experiment has demonstrated that not only a discrimination learning set but also a discrimination reversal learning set can be developed from experience with repetitive cycles of 12 problems constructed from four stimulus objects. These facts are important when it is considered that training on an extended series of discrimination learning problems fails to develop the reversal learning set even though the discrimination learning set was well established (**3, 6**). An interpretation of the differences resulting from these two experimental procedures should emphasize the importance of interproblem transfer relations in contrast to the importance of breadth of experience independent of whatever transfer that may thus arise. A concept of progressive decline of transfer is likely to play a key role in such an interpretation.

Even for the case of simple discrimination learning, evidence for the growth and decline of interproblem transfer has been presented (**8**). In the learning of a series of simple discrimination problems, interproblem transfer is not perfect, but is reduced by the variation in stimulus characteristics from one problem to the next. Transfer suppression, though adequate for efficient discrimination learning, may yet be insufficient for reversal learning. The present experiment, although it did not involve specific training on discrimination reversal, has more nearly assured the sufficiency of transfer suppression since training was continued until partial discrimination reversal problems were readily solved before complete reversal testing was attempted.

Summary

Five naive rhesus monkeys were trained on 216 simultaneous discrimination problems constructed from all possible combinations of four stimulus objects. All training was performed in an adaptation of the Wisconsin General Test Apparatus.

An analysis of interproblem transfer relations showed that partial perseveration of stimuli from one problem to the following elicits positive transfer if the stimulus plays the same role in both problems, but elicits negative transfer if its role is reversed. The magnitude of the transfer decreases with additional training. These transfer effects carry over from one problem to the succeeding problem, but not to the problem after that. Only direct reversal transfer effects persist through an intervening problem.

The *S*s were also tested for ten days on generalized discrimination and discrimination reversal learning problems. Performances on both types of problems showed consistently high efficiency of learning. These results were interpreted to indicate the significance of interproblem transfer for the formation of learning sets.

References

1. Bateson, G. Social planning and the concept of "deutero-learning." In T. H. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 121–128.
2. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
3. Harlow, H. F. Performance of catarrhine monkeys on a series of discrimination reversal problems. *J. comp. physiol. Psychol.*, 1950, **43**, 231–239.
4. Harlow, H. F. Levels of integration along the phylogenetic scale: learning aspect. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper, 1951. Pp. 121–141.
5. Magaret, A. Generalization in successful psychotherapy. *J. consult. Psychol.*, 1950, **14**, 64–70.
6. Meyer, D. R. Food deprivation and discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, **41**, 10–16.
7. Nissen, H. W., Riesen, A. H., & Nowlis, V. Delayed response and discrimination learning by chimpanzees. *J. comp. Psychol.*, 1938, **26**, 361–386.
8. Riopelle, A. J. Transfer suppression and learning sets. *J. comp. physiol. Psychol.*, 1953, **46**, 108–114.
9. Spence, K. W. Analysis of the formation of visual discrimination habits in chimpanzees. *J. comp. Psychol.*, 1937, **23**, 77–100.

(Received April 1, 1954)

Journal of Experimental Psychology
Vol. 49, No. 1, 1955

DISTRIBUTION OF PRACTICE AND THE TEMPORAL DECAY OF RESPONSE-PRODUCED STIMULI[1]

ERNST Z. ROTHKOPF

Training Aids Research Laboratory, Air Force Personnel and Training Research Center

Recent contributions to the literature on the distribution of practice have placed emphasis on two theoretical constructs. These are a negative motivational state, *reactive inhibition*, and *stimulus generalization*. The first, reactive inhibition, has been used with good success for a theoretical account of the relationship between intertrial interval and conditioning (**3, 4**), reminiscence (**2**), resistance to extinction (**4**), spontaneous recovery (**1**), and alternation behavior (**6, 9**).

Appeals to stimulus generalization as an explanatory concept in connection with experiments dealing with distribution of practice have been made chiefly with data yielded by studies involving changes in intertrial interval. Reynolds (**4**) appears to have been the first to suggest the influence of intertrial interval on the stimulus as a factor in explaining diminished resistance to extinction following changes in the intertrial period. This view was supported by Teichner (**7**) who found that rate of extinction was related to the degree of similarity between the intertrial interval employed during extinction and that used during acquisition.

Teichner interpreted his results in terms of Hull's theory (**1**). Each reaction of the organism builds up a state of negative motivation (I_R), which is temporary in character, and which dissipates as a simple decay function in time. I_R, by subtracting from the reaction potential ($_sE_R$) of the relevant response, would have as a consequence faster extinction with shorter intertrial intervals (**1**, pp. 300, 344). If one assumes that at the onset of each new trial traces of the stimulus produced by the previous response are still acting in the organism, and that the strength of this stimulation is related in some monotonic manner to the time interval which has elapsed since the last trial (**1**, p. 47), then the reinforced response becomes conditioned to those response-produced traces present at each trial for a given condition of massing. When extinction now takes place under a different condition of massing than during acquisition, S is required to respond to a stimulus complex containing traces different from those encountered before. Generalization decrement in this case implies a loss in effective habit strength ($_s\bar{H}_R$) due to the dissimilarity of the response-produced traces (s) present at an extinction trial following a given intertrial interval from those traces (δ) in S at an acquisition trial under different conditions of massing (**1**, p. 199). As the intertrial interval used during extinction becomes increasingly different from that used during acquisition, the decrement in effective habit strength ($_s\bar{H}_R$) becomes larger and has as its consequence diminished resistance to extinction. According to Hull, resistance to extinction (n) is an increasing function of the effective reaction potential ($_s\bar{E}_R$). The latter is an increasing function of $_s\bar{H}_R$, and a decreasing function of I_R. In this way, both the absolute duration of the intertrial interval used during extinction and its similarity to the intertrial interval with which S has been originally conditioned determine S's resistance to extinction.

An experiment by Teichner and Holder (**8**) demonstrated the applicability of the generalization decrement hypothesis to reminiscence phenomena. These investigators trained rats to traverse a short runway by giving them 15 reinforced trials using from zero to 3 sec. as an intertrial interval. A time period ranging from zero to 180 sec. was interpolated between Trials 15 and 16. The amount of reminiscence observed on Trial 16 was found to be a complex function of the interpolated interval: reminiscence first increasing as the interval became

[1] This experiment is part of a dissertation submitted to the Graduate School of the University of Connecticut in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author would like to thank Drs. A. M. Liberman and D. Zeaman for their interest and generous counsel.

longer, and then decreasing with still longer intervals. In terms of the theory, an increase in the interpolated rest period increases the amount of reactive inhibition (I_R) which is dissipated. This implies greater ${}_s\bar{E}_R$ following longer rest periods. However, as the rest period increases, the traces of the response-produced stimuli become increasingly unlike the zero- to 3-sec. old traces present during original conditioning. This has as its consequence increasingly smaller amounts of effective habit strength (${}_s\bar{H}_R$) and hence ${}_s\bar{E}_R$. The interaction of the two gradients accounts for the complex function reported by the authors.

The notion that the organism is operating under smaller effective habit strength when intertrial interval is changed has obvious and readily testable consequences. These are: (*a*) *Generalization Effect.* Provided I_R and Drive (D) are held constant, or are experimentally controlled, changes in intertrial interval should result in decrements in response strength. (*b*) *Relearning.* Further practice under the new intertrial interval should result in the disappearance of any response decrement attributable to a change in response-produced stimulus conditions.

In order to test these hypotheses use was made of a phenomenon observed in a recent experiment (**5**). When rats are trained to make a simple running response at one trial per day and are then changed to two trials per daily session, they show initially a decrement in running times on the second of the daily trials. Since we can safely say that the animal prior to the commencement of the two trials per day schedule has never responded to traces from a recent response, this method provides a rational technique for testing the hypotheses advanced above.[2]

Accordingly, in the present experiment a relatively strong running habit was conditioned at the rate of *one* trial every 24 hr. The animals were then tested with daily pairs of trials using various intertrial intervals according to a factorial design.

Method

Subjects

The *S*s were 19 male and 19 female albino rats of Wistar stock from the animal colony maintained by the psychological laboratory of the University of Connecticut, and 58 male albino rats of the Sprague-Dawley strain from the Rolfmaier-Holtzman Rat Company of Madison, Wisconsin. They ranged in age from 90–140 days at the beginning of experimentation. The *S*s were randomly assigned to the 16 experimental subgroups with the restriction that the various strains and sexes be as equally distributed as their number allowed.

Apparatus

Two elevated runways, each 68 in. long, 10.5 in. high, and 1.5 in. wide, and constructed of wooden blocks, comprised the main pieces of apparatus. One runway was painted flat black, the other flat white. Located at the end of each runway was a small elevated cup which contained wet Purina Laboratory Chow mash during all trials. The runways were placed on tables which stood on opposite sides of a well-lighted experimental chamber. A stool, on which *S*s were placed between trials, stood 24 in. from the starting side of each table. Two Standard Electric clocks coupled to a hand-switch were used to record latency and running-time measure.

Procedure

Habituation.—One week prior to the experiment proper all *S*s were placed on a 24-hr. feeding schedule, and were frequently handled and habituated to feeding in various places such as table tops, strange cages, feeding platforms, and the like.

Phase A.—Following habituation *all* *S*s received ten reinforced trials at the rate of one trial per day. Latency and running time were

[2] It has been assumed that the stimulus continuum associated with intertrial interval is related to the decay of response-produced stimulation in time. Since the trace is presumably the result of muscular contraction, it is possible to infer the characteristic of a given trace from the kind of response made and the time elapsed since that response. Reynolds (**4**), however, has pointed out the possibility of traces from several massed responses summating with each other. This factor would obviously complicate the inference of trace magnitude. Using only one pair of trials per day avoids this complication because all relevant traces are allowed to dissipate completely between days.

measured on each trial. The latency clock was started as *S* was placed on the runway and stopped when the base of *S*'s tail passed over a marker located 10 in. from the starting edge. Running time was recorded as the interval between the stopping of the latency clock and the time *S* reached the food cup. In all trials *S* was allowed to feed for 5 sec. after reaching the goal.

Phase B.—On Day 11 the *S*s were divided into four groups. Group A received one reinforced trial which was followed as soon as possible by another reinforced trial. Actually, this time interval was 2–5 sec., which was the time required to place *S* on the waiting platform, record the data on magnetic tape, reset the clocks, and replace *S* on the runway. This time interval will henceforth be referred to as the "5-sec." interval. Groups B, C, and D also received two reinforced trials but 15, 30, and 120 sec., respectively, were interpolated between the first and second trials. The *S*s remained on a nearby waiting platform throughout these intertrial periods. This procedure was repeated at 24-hr. intervals for six days in all.

Phase C.—Each one of the groups described above was divided into four equal subgroups which represented the four intertrial intervals used in this experiment. For example, Group A which had operated on a 5-sec. interval was now split into A-1, A-2, A-3, A-4. Each *S* of Group A-1 still received two reinforced trials per day 5 sec. apart, *S*s of A-2 received two reinforced trials 15 sec. apart, and so on. One-fourth of the experimental population, namely, A-1, B-2, C-3, and D-4, was still operating under the original training conditions, while the rest had new intertrial intervals. Phase C, although intended to last six days, consisted of five reinforced pairs of trials over a period of five days because of *E*'s error.

Phase D.—Groups A-1, B-2, C-3, D-4 continued receiving reinforced pairs of trials using the original intertrial interval at the rate of one pair of trials per day for a period of 8 days, giving each of these subgroups a total of 19 reinforced pairs of trials.

All other *S*s received six additional pairs of trials under each of the two intertrial intervals with which they had as yet no experience.

For the purposes of a planned study not related to the present theoretical problem, half of the trials of each phase (except the five-day Phase C) for each *S* were on the white runway and the other half on the black. Both trials on any one day were in all cases on the same runway.

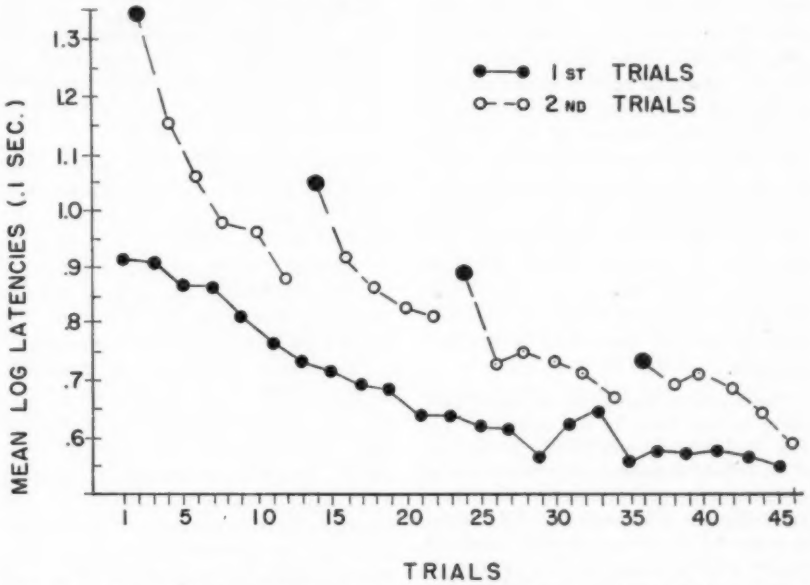


FIG. 1. Mean log latency of the two daily trials for those *S*s who received training under all four of the intertrial intervals used in this experiment. The heavily accentuated circles represent measures obtained on the second trials of days on which a general change in intertrial interval was effected.

Results and Discussion

In order to test the hypothesis that *changes* in intertrial interval produce reductions in effective habit strength it is necessary to provide a set of conditions in which the time afforded for the dissipation of reactive inhibition is held constant. This is the case for the data presented in Fig. 1. These were obtained from *S*s (all groups except A-1, B-2, C-3, D-4 which contribute only to the first 12 trials) that received blocks of paired trials in which the intertrial interval varied from block to block. Each measure shown in Fig. 1 beyond Trial 12 has been obtained, therefore, from an *N* of 72 in which subgroups of 18 *S*s were operating under 5-, 15-, 30-, and 120-sec. intervals, respectively. The individual latencies in .1-sec. units have been transformed into logarithms in order to normalize them. The obtained function indicates that following a change in intertrial interval (the heavy circles in Fig. 1), in which each *S* is switched to a different condition of massing *but in which the total number of 5-, 15-, 30-, and 120-sec.* S*s remains the same*, there is a weakening in response strength on the second trial of the day. There is recovery from this weakening as a function of practice with the new interval. The initial decrement following a general change in intertrial interval grows less with successive changes.

These results cannot be accounted for solely on the basis of the construct I_R. The total time allowed for the dissipation of I_R is held constant from trial to trial in the experimental population contributing to Fig. 1. The presence of I_R might be expected to reduce response strength on the second trials but this decrement should remain constant throughout the course of the experiment. The decrement shown in Fig. 1 does not remain constant in this case.

The data can be accounted for, however, if one considers the decrement observed on the second trials as being due both to the effect of I_R *and* to the weakening of the second response which results from the fact that the stimulus complex which elicited this response was only partially identical to that which obtained at a trial following a 24-hr. rest. On the second trial of each day *S* encountered the traces of response-produced stimulation of the first trial. This produced a generalization decrement. During reinforced practice with these traces, habit strength was built up for that component of the stimulus, and response strength for the second trial increased. When *S* was switched to a different intertrial interval, the traces now encountered at the second trial of the daily pair were different from those which had prevailed under the previous condition and a generalization decrement again was observable. As *S* gained experience with more intertrial intervals, the effective habit strength for those trace conditions which had as yet not been encountered rose through generalization processes. This accounted for reduction in the initial postchange decrements as a function of repeated changes (the heavy circles in Fig. 1).

If only the eight data points supplied by the second-trial measures of the first and last day of each of the four phases are considered, it may be noted that the hypotheses elaborated in the preceding paragraph require that (*a*) the response time of the second daily trial at the beginning of each phase become successively smaller with each successive phase; (*b*) the response time of the second daily trial on the last day of each phase become successively smaller with each successive phase; (*c*) the response time of the second trial of the first day of any phase be larger than the second-trial measure of the last day of that block; (*d*) the second-trial response time of the last day of any phase be smaller than the second-trial

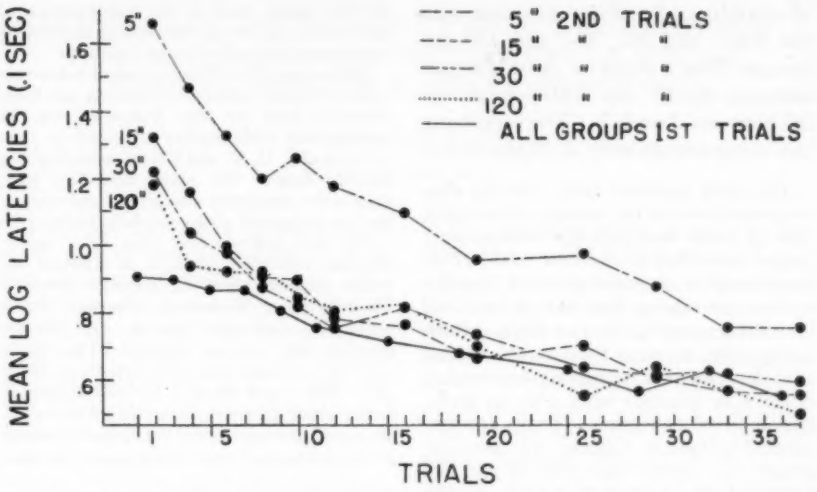


FIG. 2. Mean log latencies of first and second trials of the daily pair of trials for *Ss* operating under only one intertrial interval throughout the course of the experiment. All points beyond Trial 12 are the mean values of two consecutive days.

response measure of the first day of the following phase. Only five of the 40,320 ways in which eight measures may be ordinally arranged will satisfy all four of these conditions. This suggests that, if a relatively stringent criterion is used, results such as those shown in Fig. 1 may occur by chance with a probability of only .0002. Similar results were obtained with the running-time data.[3]

The data presented in Fig. 2 similarly required an interpretation in terms of I_R and generalization decrement. The functions shown here were obtained from *Ss* which, following one trial per day, were switched to one pair of trials per day that were separated by the various intertrial intervals used in this experiment. They differed from the population shown in Fig. 1 in that each *S* was operating under only *one* intertrial-interval condition throughout the course of the experiment. An analysis of covariance, in which the mean log latencies of the various groups on Trial 2 have been adjusted on the basis of performance on Trial 1, yielded an F of 9.49 between groups. This F, with 3 and 91 df, is significant beyond the .01 level of confidence. Results of t tests among the various groups indicated that all differences except those between the 15- and 120-sec. groups, and the 30- and 120-sec. groups were beyond the .05 level of confidence. The p value for the differences between the means of the 15- and 120-sec. group was between .1 and .2. The null hypothesis could not be rejected for the comparison of the 30- and 120-sec. groups. When the running-time measures were subjected to a similar analysis an F of 8.19 was obtained ($p < .01$). The t tests were significant at the .01 level

[3] The running-time measures as well as the data on which the graphs and statistical materials of this paper are based have been deposited with the American Documentation Institute. Order Document No. 4413, remitting $1.25 for microfilm or $1.25 for photocopies.

of confidence for differences between the 5-sec. and 15-, 30-, and 120-sec. groups. The p value of the difference between the 15- and 120- sec. groups fell between .1 and .2. The remaining two comparisons were at chance level.

The data indicate fairly clearly that response time on the second of the daily pair of trials was initially considerably longer than that of the first trial. This weakening in response strength appears to be a decreasing function of the time interval separating the two trials. Both latency and running time on the second trial decrease in a negatively accelerated manner as practice with a given intertrial interval continues. There are theoretical reasons to suppose that any decrement not further reducible by practice represents the contribution of I_R to the postchange loss in response strength. In view of the fact that after 19 reinforced pairs of trials only the 5-sec. groups shows a decrement on the second trial of the daily pair, it appears that I_R effects in this situation are slight, and that the initial postchange losses in response strength represent a good estimate of the generalization effect. These initial decrements when plotted against change in intertrial interval appear to approximate a simple decay function. This is to be expected when it is assumed that the magnitude of the response-produced trace is a decreasing, negatively accelerated function of intertrial interval, and that changes in ${}_{S}\bar{H}_R$ are linearly related to changes in the magnitude of the trace. Further research is required to determine the exact nature of the relationship between generalization decrement and changes in intertrial interval.

Summary

The present experiment provides an empirical test for two hypotheses based on an elaboration of Hullian behavior theory: (*a*) *Generalization Effect.* Providing I_R and D are held constant or are experimentally controlled, changes in intertrial interval should result in decrements in response strength. (*b*) *Relearning.* Further practice under the new intertrial interval should result in the disappearance of any response decrement attributable to a change in response-produced stimulus conditions.

Albino rats ($N = 96$) were trained to traverse a short elevated runway by receiving one food-reinforced trial per day. Following this, *Ss* received two trials per day separated by time intervals of 5, 15, 30, and 120 sec. according to a factorial design. The major dependent variables of the experiment were latency and running time on the second of the two daily trials.

The data indicate that when I_R is experimentally controlled, changes in intertrial interval produce losses in response strength. These response decrements disappear in a negatively accelerated manner with further practice under the new interval. The experiment thus confirms both of the hypotheses under test. The results obtained are held consonant to the notion that response-produced traces are an important component of the stimulus complex to which *S* learns to respond in massed practice.

References

1. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
2. Kimble, G. A., & Horenstein, B. Reminiscence in motor learning as a function of length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239–244.
3. Reynolds, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, **35**, 15–30.
4. Reynolds, B. Extinction of trace conditioned responses as a function of the spacing of trials during the acquisition and extinction series. *J. exp. Psychol.*, 1945, **35**, 81–95.
5. Rothkopf, E. Z., & Zeaman, D. Some stimulus controls of alternation behavior. *J. Psychol.*, 1952, **34**, 235–255.
6. Solomon, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, **45**, 1–40.
7. Teichner, W. H. Experimental extinction as a function of the intertrial intervals during conditioning and extinction. *J. exp. Psychol.*, 1952, **44**, 170–178.
8. Teichner, W. H., & Holder, E. Reminiscence as a function of change in the intertrial interval. *J. exp. Psychol.*, 1952, **44**, 347–351.
9. Zeaman, D., & House, B. J. The growth and decay of reactive inhibition as measured by alternation behavior. *J. exp. Psychol.*, 1951, **41**, 177–186.

(Received March 31, 1954)

Journal of Experimental Psychology
Vol. 49, No. 1, 1955

LEARNING AND PERFORMANCE ON A KEY-PRESSING TASK AS FUNCTION OF THE DEGREE OF SPATIAL STIMULUS-RESPONSE CORRESPONDENCE[1]

ROBERT E. MORIN[2] AND DAVID A. GRANT

University of Wisconsin

The chief aim of the present experiment was to determine how learning and performance measures are influenced by changes in the degree of spatial correspondence between the stimulus elements and the response elements of a motor task. The problem originated in connection with the design of display-control systems. It led to considerations regarding the scaling of difficulty of perceptual motor tasks, a more general problem.

If display elements and response elements are arrayed in horizontal rows from left to right, the work of Fitts and Seeger (**2**) suggests that the most efficient arrangement would be direct spatial display-control correspondence where the left-most control element would be paired with the left-most stimulus element, the next control with the next stimulus, etc. The system designer must, however, consider several objectives, and often one objective will be met at the expense of another. Added weight, cost, or complexity of equipment must be weighed against possible losses in performance efficiency when other than direct correspondence relations are used, and this experiment was designed to provide a quantitative measure of performance degradation under quantitatively scaled departures from direct correspondence.

General consideration of the stimulus-response correspondence problem revealed that certain attacks on it might yield a means of obtaining a series of homogeneous psychomotor tasks that would differ in difficulty. This would be a significant achievement, because one difficulty in the orderly investigation of motor skill has been the lack of materials of scaled difficulty such as those which exist for rote and verbal learning (**5**).

In the present study, the display consisted of parallel, horizontal rows of eight stimulus lights and eight response information lights. The controls consisted of eight response buttons in a row parallel to the lights. The same general task was carried out by *S* in all instances, but the difficulty was altered by the degree to which display-control connections departed from direct stimulus-response correspondence. In general, *S*'s task was to match the pattern of illuminated stimulus lights by depressing those keys which lighted response information lights directly below the stimulus lights. Varying degrees of stimulus-response correspondence were obtained by changing the functional connections between stimulus lights and response keys without any modification of the external physical properties of the equipment seen by *S*. Stimulus lights and response keys were numbered 1, 2, . . ., 8, from left to right. In direct correspondence, Lights 1, 2, . . ., 8 would be wired to Keys 1, 2, . . ., 8, respectively. In a departure from direct correspondence, Lights 1, 2, . . ., 8 might, for example, be wired to Keys 1, 2, 4, 5, 3, 6, 8, 7, respectively. There were several ways in which the degree of correspondence could have been specified. The most feasible seemed to be a measure of rank correlation computed between the two series of numbers representing the stimulus lights and the response keys that were, respectively, connected to them. This index had the advantage of specifying the

[1] This research was supported in part by the United States Air Force under Contract Number AF 18(600)-54 monitored by the Aero-Medical Laboratory of the Wright Air Development Center, Wright-Patterson Air Force Base, Ohio. An extended version of the results was published in WADC Technical Report 53-292.

[2] Now at the University of Texas.

correspondence system (the independent variable) with a single number.

An attempt was made to predict the amount of performance degradation to be expected with different functional pairings of stimulus lights and response keys. It was first assumed that the habits associated with direct display-control correspondence should have been the ones most thoroughly learned through daily experience. In other words, a rank correlation of +1.00 should produce best performance. Nondirect correspondence habits should have been less well-learned and also might be interfered with by the pre-experiment experiences of direct spatial stimulus-response correspondence. Three predictions seemed plausible with respect to the amount of performance degradation: (*a*) Assuming that the direct correspondence response would be the one most highly learned in pre-experimental experience, the difficulty of learning other responses to this same stimulus should be a decreasing monotonic function of response similarity (**3, 4, 9**). If the over-all index of stimulus-response correspondence can be used to indicate response similarity, then performance should be best for a correlation of +1.00 and show progressively more degradation as the correlation approaches −1.00. In the simplest case, this would be a linear factor. (*b*) A second alternative is that similarity and difficulty may be related to the absolute value of r. This consideration leads to the prediction that difficulty might be approximated by a quadratic or some other even function of stimulus-response correspondence, with the greatest difficulty coming at a correlation of 0.00. This alternative has some intuitive appeal. It recognizes that decoding may be easier for *S* when there is order in the response set, even though that order may be opposite to that of the stimulus set. (*c*) The third, and in many ways the most attractive, hypothesis is a combination of *a* and *b* which would predict difficulty to be the net outcome of a monotonic increasing function and a superimposed quadratic function of r.

Method

Apparatus.—The Multiple Serial Discrimeter (MSD) was used in the present study. The MSD is best described in terms of its five basic, independent units: (*a*) response keyboard, (*b*) display unit, (*c*) stimulus programming unit, (*d*) control unit, and (*e*) graphic recorder.

a. The response keyboard is similar to ones which have been used by other investigators (**7**). It consisted of eight Lucite keys mounted like typewriter keys on the top of a flat metal box. The four left-hand keys were operated by the four fingers of the left hand, and the four right-hand keys by the fingers of the right hand. The keys were 1 in. apart, except that the middle two were separated by 3 in. A metal shield above the keys prevented *S* from seeing his hands while they operated the controls.

b. The stimulus display consisted of two parallel, horizontal rows of jeweled lamps mounted on a vertical panel. The top row consisted of eight red jeweled stimulus lamps, 1 in. apart, except that the middle two were 2.5 in. apart. A green response light was placed 1.5 in. directly below each stimulus light. The stimulus lights were operated by the stimulus programming unit that is described below, and the response lights were operated by the keys of the response keyboard described above. When the red lights were illuminated, *S* had to light the corresponding green lights. Under direct correspondence, the response key farthest to the left operated the green light farthest to the left on the display panel, and each succeeding key operated each succeeding light. A switching board on the control unit permitted modification of this direct correspondence arrangement so that any degree of noncorrespondence could be obtained. The green information lights were not a necessary element to the task, as they could be switched out of the circuit, which was done at one point in the study.

c. The stimulus programming unit consisted of a pair of Western Union telegraph tape transmitters and associated relays. All possible combinations of stimulus lights were obtainable in any sequence by punching the proper holes in the tape by means of a Western Union tape perforator.

d. The control unit consisted of a set of

relays for comparing stimulus and response elements, and an electronic timer. The MSD is built so as to permit either self-paced or forced-paced presentation of stimulus elements. In the present study, the rate of stimulus presentation depended upon *S*'s responses. A new pattern of lights was displayed as soon as the previous pattern had been correctly matched. For a response to be correct and a new pattern to come on, the correct keys had to be down at the same time, with none of the incorrect keys down at that time.

e. A 20-pen Esterline-Angus graphic operations recorder was used to record details of errors and correct responses. Each of the eight relays which controlled the eight stimulus lights also operated a pen on the recorder and, in the same manner, the eight response key relays were associated with pens. From this record it was possible to obtain time per pattern, response latencies, total responses, and errors.

Experimental design.—There are eight factorial or 40,320 possible arrangements or permutations of response keys relative to the fixed stimulus lights. These possible arrangements may be scaled by use of Kendall's τ (6). Tau is a measure of rank correlation and may be calculated from the following formula: $\tau = \frac{2S}{N(N-1)}$ where N equals the number of items in either variate. The value for S may be obtained as follows: (*a*) First rearrange one variate in an ascending order, then (*b*) assign to each of the $\frac{N(N-1)}{2}$ possible combinations of two items of the other variate a score of + 1, if the items in the combination are in ascending order, and a score of − 1, if the items are in descending order, then (*c*) S = the sum of these scores. The distribution of S and the associated distribution of τ for an N of 8 are given by Kendall (6, p. 38).

The key-light arrangements were determined by first arbitrarily selecting nine levels of τ, four positive, four negative, and zero. At each of these levels three key-light arrangements were randomly selected. Since there was only one ordering of the variates which gave a correlation of plus or minus 1.00, only 23 *different* arrangements were used in all. Nine *S*s were tested at each level of τ, three with each key-light arrangement.

Procedure.—Two stimulus lights were presented on each trial to form a stimulus pattern. A total of 275 two-light patterns, subdivided into 11 blocks of 25 patterns each, were presented in the experimental session. During the first ten blocks, the key-light arrangements used gave τ's of − 1.00, − 0.86, − 0.57, − 0.29, 0.00, + 0.29, + 0.57, + 0.86, and + 1.00, respectively, for the nine groups. For the first nine blocks of patterns, the green information lights operated whenever a response key was depressed, but during the tenth block the green lights were switched off. On the eleventh block of patterns, the green response lights were operating again, and for all groups the keys were put into direct display-control correspondence with the stimulus lights; i.e., $\tau = + 1.00$. Between successive blocks of patterns there was a 1-min. rest period. During the first eight rest intervals, *S*s described eight pictures. They talked about each picture for the whole minute to prevent rehearsal of the key-light connections.

During the instructions, the apparatus was demonstrated with the keys and lights in direct correspondence. It was emphasized to all *S*s, however, that they should not expect this relationship during the test session. They were instructed to try to learn the key-light connections while working at matching patterns. No mention was made of a later series without the green lights (Block 10) or with the keys and lights in direct correspondence (Block 11). The instructions for these changes in test administration were given in the rest periods preceding them. The *S*s were told that each pattern would stay on until they matched it correctly, and that each new pattern should be matched as quickly as possible. Nothing was said about attempting to avoid errors.

Two measures were used. The first was mean time per pattern. The second was mean number of key presses per pattern. The latter measure was a reflection of the number of errors made. A minimum of two key presses per pattern, or 50 key presses for the 25 two-light patterns, was required. (In practice, it was possible to complete a set of 25 patterns in a few less presses since, if one light of a pattern was the same as one light of the previous pattern, one key did not have to be released before matching the second pattern. Most *S*s did release both keys. The contribution of this error to the total key-press measure was small enough to be disregarded, so that all presses over 50 per block, or over two per pattern, were considered as errors.)

Subjects.—The *S*s were 81 male students at the University of Wisconsin who were selected from a card file of students who had volunteered to serve as paid subjects. The *S*s were unsystematically assigned to the nine experimental groups at the convenience of *S* and *E*. Some *S*s had been employed in other types of experiments in the laboratory, but none of them had worked on the MSD.

Results

The first analysis was of the homogeneity with respect to performance

degradation of the three permutations or key-light arrangements within each level of τ (excepting for a τ of plus or minus 1.00). Only the arrangements for $\tau = .86$ differed significantly ($p < .01$). But since the probability of at least one such significant F among seven independent analyses is .0679, it was assumed that the one obtained might be a Type I error, and it appeared, therefore, that the variability among arrangements within a constant τ might be attributed to individual differences.

Further analyses were in terms of the mean time per pattern in seconds and the mean number of key presses per pattern at different stages of the experiment and for different levels of τ. All of the data in terms of mean time per pattern are given in Fig. 1 where mean time per pattern is plotted against successive blocks of practice trials with the value of τ as the parameter. The data from Blocks 10 and 11 have been added to the learning curves to complete the picture. To simplify the graphical presentation, the acquisition curves for the negative correlation groups are presented in the left-hand portion of the figure, and the zero and positive correlation groups are given in the right-hand portion of the figure. During Blocks 1–9, the mean time per pattern decreases regularly for each group with very little intermingling of the means from group to group. The regularity of these data based on only nine Ss per point is noteworthy. The absolute amount of improvement in performance can be seen to be a positive function of the initial difficulty for each level of τ. With the possible exception of the curve for $\tau = 1.00$, all curves appear to be dropping after

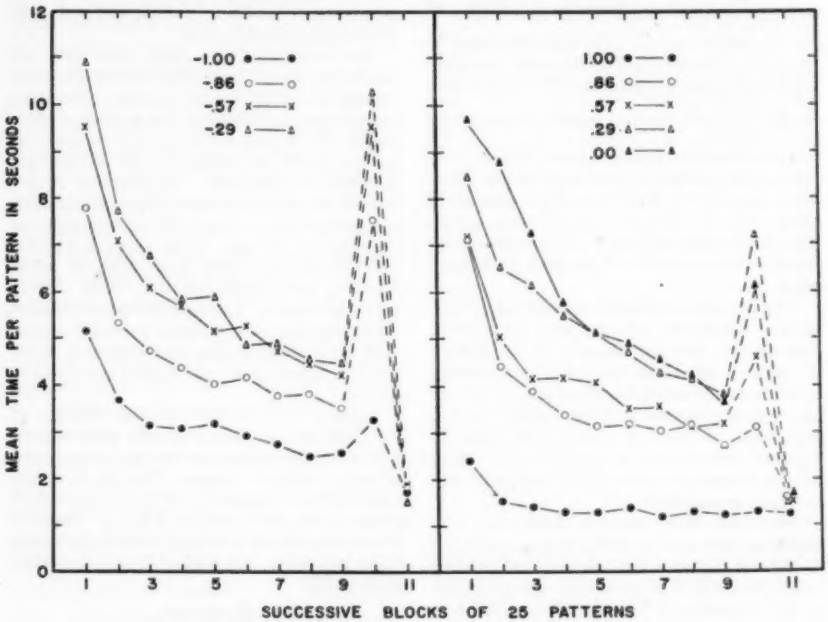


FIG. 1. Speed of performance acquisition curves with τ as the parameter.

nine blocks of practice so that it is impossible to compare the groups in terms of asymptotic performance. When the supplementary information feedback provided by the green response lights was eliminated in Block 10, all groups except that for $\tau = 1.00$ showed a marked increase in mean time per pattern. Especially large increases were noted in the $-.86$, $-.57$, and $-.29$ groups. The greatest increment in mean time per pattern was observed in the $-.29$ group where *Ss* took an average of 10.34 sec. per pattern on Block 10, or 5.83 sec. more per pattern than they required for Block 9. In general, the amount of loss in performance efficiency seemed to be directly related to the difficulty during Blocks 1–9. It may be concluded, therefore, that the green lights provided important visual information and that their importance was directly related to the difficulty of the task or the lack of correspondence between display and control elements of the task. The last set of points in Fig. 1 give the data for Block 11. On these final 25 patterns, all *Ss* practiced with the response keys and stimulus lights in direct correspondence. It will be noted that the data are markedly homogeneous on Block 11 despite differences in earlier training.

A similar set of curves could be obtained in terms of mean key presses per pattern. Since the essential relationships were identical with those in Fig. 1, the mean key presses per pattern data will not be presented separately.

At each stage of learning, it would be possible to replot the data of Fig. 1 as a function of τ. In Fig. 2 this has been done for the average performances on Blocks 1–9. The response speed is greatest for $\tau = +1.00$, next best is $\tau = -1.00$, and poorest for

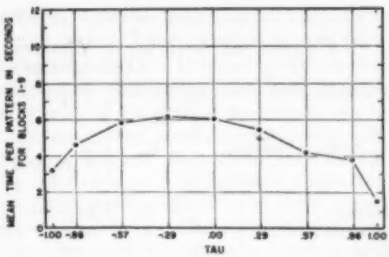


FIG. 2. Speed of performance, averaged over Blocks 1–9, as a function of τ.

$\tau = -.29$. There are no inversions, and two general characteristics of the curve in Fig. 2 are apparent: (*a*) the curve appears much like a parabola with (*b*) its axis rotated from the vertical. These two characteristics suggest the presence of linear and quadratic components whose significance can be tested by means of orthogonal polynomials (1). This technique enables one to determine details of the functional relation that exists between performance and the scaled variable τ. It is possible to test independently whether a curve has significant linear, quadratic, cubic, etc. components. Strictly speaking, the analysis by orthogonal polynomials is proper only when the intervals of the scaled variable are equal. In this study, the two outer intervals were only one-half the size of the six inner intervals, which would tend to reduce somewhat the sensitivity of the statistical test; but this source of downward bias was negligible, particularly with lower degree polynomials.

The *F*'s for an analysis in terms of orthogonal polynomials are given in Table 1, Column A, for the mean time per pattern on Blocks 1–9. Each of the curves being analyzed in Table 1 has nine data points or 8 *df* between the means which define the points. These have been subdivided into 1 *df* for each term from linear to quintic

TABLE 1

F's from Six Analyses of Variance, Evaluating Performance as a Function of τ by Orthogonal Polynomials

Source of Variation	*df*	Mean Time per Pattern			Mean Key Presses per Pattern		
		A. Blocks 1–9	B. Block 10	C. Block 11	D. Blocks 1–9	E. Block 10	F. Block 11
Linear	1	23.89***	17.66***	10.93**	17.20***	7.26**	3.44
Quadratic	1	64.59***	34.69***	1.69	88.57***	13.25***	.38
Cubic	1	.63	8.32**	.32	1.61	7.92**	2.14
Quartic	1	.25	1.05	3.00	.23	.93	2.04
Quintic	1	2.20	.09	.43	1.05	.07	.16
Higher order	3	.41	1.63	.14	1.10	1.67	1.44
Error variance (Within τ M.S.)	72	1.131	10.067	.104	.938	24.424	.039

** Significant at .01 level.
*** Significant at .001 level.

polynomials and 3 *df* for the pooled sixth-, seventh-, and eighth-order polynomials. The error term with 72 *df* is based on the variability of *Ss* within levels of τ. At the head of each column of *F*'s in Table 1 are given the measures used in the analysis and the experimental period analyzed. As was suggested by the graphic presentation of data in Fig. 1, the curve has significant quadratic and linear components but none of the higher-order components was significant. A similar plot of mean key presses per pattern for Blocks 1–9 looked very similar to Fig. 1, and the orthogonal polynomial analysis of these data are given in Column D of Table 1 which shows a significant linear and quadratic term as was also observed for the data of Fig. 2.

When mean time per pattern on Block 10 was plotted against τ the curve was very similar in appearance to that of Fig. 2 except that it went higher (to an average of 10.34 sec. per pattern for $\tau = -.29$), especially for the minus correlations between zero and -1.00. The curve was consequently more skewed than that of Fig. 2, and considerable skewness remained even when the axis of the curve was rotated so as to effectively eliminate the linear component. There was also some sign of inflection for positive values of τ. Since the orthogonal polynomial analysis uses a priori orthogonal coefficients, the remaining skewness and inflection could introduce a significant third-order component into the analysis. This was observed as is shown in Column B of Table 1 where the cubic component is significant in addition to the linear and quadratic terms. The mean key presses per pattern analyzed in Column E of Table 1 showed the same result for the same reason. The significance of the cubic component shows that the absence of the green response lights had a markedly differential effect on the different groups in the experiment.

In the final block of trials, with the keys and lights in direct correspondence for all groups, plotting mean time per pattern against τ in early practice yields a curve which stays at 1.7–1.8 sec. with minor fluctuations and a slight downward trend over all values of the independent variable except for a drop of .4 sec. at $\tau = +1.00$.

The linear component in Column C of Table 1 was significant, but an analysis (not reported in Table 1) showed that if this point were omitted the linear component was not significant. In assaying the amount of positive transfer effect it is interesting to note that the performance on Block 11 of groups with $\tau \neq 1.00$ on Blocks 1–10 corresponded to the Block 2 average of the group that was given continuous practice with $\tau = 1.00$.[3]

Data for mean key presses per pattern on Block 11 were similar except that the drop for $\tau = 1.00$ did not occur. None of the orthogonal components of the variance of mean key presses per pattern was significant on Block 11. There was, therefore, complete positive transfer on error scores in all groups, and in all groups errors were uniformly low. This indicates direct correspondence habits have a high resistance to overt intrusions (errors) from nondirect correspondence training.

Discussion

Throughout this experiment the best performance was obtained with the response keys and the stimulus lights in direct correspondence. This was not surprising. A different result would hardly have been expected. It was not sufficient, however, to know that a correlation of +1.00 between stimulus and response elements produced superior performance. More important was information on the magnitude of differences in performance, their permanence, and how they were related to the scaled variable, τ. For the most part, clear answers to these three questions were obtained from the data of the first nine blocks of patterns.

The results indicate that direct correspondence performance was very much superior to all other conditions studied. Changes from direct correspondence multiplied the average time per unit of work at different stages of practice by a factor of 2 to 5 and increased errors by a factor of 4 to 15. It is important to emphasize in this connection how greatly even slight changes from direct correspondence may lower performance efficiency. The τ of .86 involves interchanging only two pairs of lights or displacing one light two positions away from direct correspondence. These simple interchanges ought to be easily learned, and little, if any, detrimental effect on performance might be anticipated. Figure 1 shows that this anticipation is far from the truth. If the curves for $\tau = +1.00$ and $\tau = .86$ are compared, it is observed that the absolute difference between the groups is diminshed only slightly by practice. Even after nine blocks of practice, the .86 groups were taking more than twice as much time per unit of work as the direct correspondence group. All τ groups might reach asymptotes at about the same level with very extended practice, but the likelihood seems small. It is indicated, therefore, that practice will compensate very little for loss of correspondence or compatibility. Additional weight, complexity, and cost of equipment may, therefore, have to be accepted where speed and errorless performance are critical in this type of task.

Performance proved to be related to τ by regular and rather simple functions that were in line with Prediction c; i.e., both the algebraic value and the absolute value of τ played a role in determining difficulty as shown by the significance, respectively, of the linear and quadratic components of the performance measures. It is evident that even though every individual S-R connection is displaced maximally, a complete reversal of display-control relationships ($\tau = -1.00$) yields better performance than do the more irregular rearrangements ($-.86 \leq \tau \leq +.86$). This quadratic component emphasizes the fact that S responds readily to orderliness, direct or reversed (high positive or negative values of τ). But, in addition, it was noted that the linear component of the performance functions was significant, which meant that for each negative τ the performance

was poorer than that observed at a positive τ of the same absolute value. It is believed, therefore, that a double process of decoding was required of *S* for negative values (except -1.00) of τ, especially on the high values. The *S* apparently responded to the reversal, and also learned the departures from the reversed display-control relationships.

Final observations may be made on the data. First, the data of Fig. 1 permit the wide choice of levels of task difficulty for a homogeneous psychomotor performance. An initial difficulty ratio of five to one is provided by choosing $-.29$ with $+1.00$ arrangements, and a final difficulty ratio of three to one for Block 9 is provided by a selection of the same two groups. Intermediate ranges are provided by the selection of other τ values. Thus a wide range of task difficulty can be obtained by the selection of appropriate levels of τ. Second, the data of Block 11 show that it is very difficult to interfere with direct correspondence responses. All groups transferred to direct correspondence in Block 11 after training at one of the nine levels of τ. On the assumption of pre-experimental learning of direct stimulus-response relationships, this is essentially the paradigm for retroactive inhibition since new responses were connected to the old stimuli in the "interpolated" learning (**9**). Instead of inhibition, facilitation was found. This was probably a result of (*a*) the relatively high level of skill the *S*s brought to the task for direct correspondence operations (**8, 10, 11**) and (*b*) the fact that the first ten blocks of trials offered an opportunity for response differentiation. Stated simply, *S* learned the "feel" of the keyboard during the training period. Although it was difficult, in Blocks 1–10, to interfere with the "retention test" of direct correspondence performance on Block 11, it is suspected that it would be very easy to interfere with trained performance on one of the more difficult τ conditions.

It was observed that *S*s were usually able to schematize perfectly on paper the stimulus-light response-key connections which they had practiced. In this sense, they had learned, but they could not decode rapidly enough for fast performance. They were dependent upon the green response information lights which were available as response feedback on the first nine trials of the experiment. Turning off the green lights on Block 10 removed this supporting feedback and resulted in serious degradation of the performance. An interesting question is whether a partial information (reinforcement) schedule would have reduced or eliminated dependence on the information lights. This last consideration is part of the general problem of the use in a training situation of cues which are not present in the criterion or tactical performance.

Summary

Eighty-one *S*s in nine groups practiced key-pressing responses to light stimuli under nine different degrees of spatial stimulus-response correspondence between lights and keys. The degree of correspondence was specified by Kendall's τ, a measure of rank correlation. The basic findings were:

1. Performance was uniformly superior when keys and stimuli lights were in direct correspondence ($\tau = 1.00$).
2. The marked degradation of performance in other conditions proved to be related to τ by regular and rather simple functions. The function relating mean time per pattern to τ had significant linear and quadratic components. These were hypothesized to be related, respectively, to factors defined by the algebraic and absolute values of τ.
3. Scaling of correspondence by τ provided a set of homogeneous psychomotor tasks of scaled difficulty with an initial difficulty range in time scores of five to one and a final range of three to one.
4. Nondirect correspondence groups showed a high dependence on a set of response information feedback lights.
5. In a final test in which all groups transferred to direct correspondence practice, facilitation rather than interference was found.

References

1. Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural and medical research.* (2nd Ed.) Edinburgh: Oliver & Boyd, 1943.

2. Fitts, P. M., & Seeger, C. M. S-R compatibility: spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.
3. Gagné, R. M., Baker, K. E., & Foster, H. On the relation between similarity and transfer of training in the learning of discriminative motor tasks. *Psychol. Rev.*, 1950, **57**, 67–79.
4. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.
5. Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 517–567.
6. Kendall, M. G. *Rank correlation methods*. London: Griffin, 1948.
7. Klemmer, E. T., & Muller, P. F. The rate of handling information: key pressing responses to light patterns. Washington: Human Factors Operations Research Laboratories, Bolling Air Force Base, 1953. (*HFORL Memo Rep.* No. 34.)
8. Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.
9. Osgood, C. E. The similarity paradox in human learning: a resolution. *Psychol. Rev.*, 1949, **56**, 132–143.
10. Sears, R. R., & Hovland, C. D. Experiments on motor conflict. II. Determination of mode of resolution by comparative strengths of conflicting responses. *J. exp. Psychol.*, 1941, **28**, 280–286.
11. Siipola, E. M., & Israel, H. E. Habit-interference as dependent upon stage of training. *Amer. J. Psychol.*, 1933, **45**, 205–227.

(Received April 1, 1954)

Journal of Experimental Psychology
Vol. 49, No. 1, 1955

ROTARY PURSUIT PERFORMANCE UNDER ALTERNATE CONDITIONS OF DISTRIBUTED AND MASSED PRACTICE

M. RAY DENNY, NORMAN FRISBEY, AND JOHN WEAVER, JR.

Michigan State College

The present study is concerned with three constructs which have been postulated to account for rotary pursuit phenomena; namely, temporary inhibition (reactive inhibition, temporary work decrement), conditioned inhibition, and warm up or set (**1, 2, 3, 5, 6, 7**). The study closely parallels an investigation by Weaver (**10**), except that Weaver used a disk which rotated in an unconventional counterclockwise direction. Weaver's results indicated that the effects of massing vs. spacing were not permanent: Temporary inhibition built up rapidly to a maximal value and dissipated just as rapidly; conditioned inhibition which built up during continuous practice extinguished when the conditions were shifted to spaced practice.

In view of his findings we made the following predictions: (*H-1*) Irrespective of the previous conditions of practice, groups practicing under the same conditions will converge to the same performance level. (*H-2*) It will be possible to derive an index of conditioned inhibition which will be an increasing function of the number of massed trials and will extinguish during spaced or distributed practice as a familiar decay function. (*H-3*) Given three practice periods separated by rests and the variable of massed vs. distributed practice, the group working under massed-distributed-massed (M–D–M) conditions will show a more immediate decline in performance in the third practice period than any other combination of conditions. This last prediction was made because *reconditioning* of previously extinguished conditioned inhibition is involved in the M–D–M group in the third period and reconditioning is typically rapid.

With performance measured in 10-sec. units, as well as 30-sec. units, it was possible on the basis of the factor of warm up or set to make the following prediction with respect to the trial directly following the first long rest period: (*H-4*) For the first postrest 30-sec. trial, the score on the third 10-sec. unit will be higher than on the first 10-sec. unit, whereas the opposite relationship will be true for the prerest 30-sec. trials.

METHOD

Subjects.—A total of 76 students from psychology classes at Michigan State College served as *Ss*. No *S* had had prior pursuit rotor experience. Twelve of the *Ss'* records were omitted in order to match the groups, leaving 64 *Ss*, 36 of whom were female. Mean age was 20.6 yr. The first 48 *Ss* were randomly assigned to one of eight subgroups while the remainder were assigned, for the purpose of matching, on the basis of the first 30 sec. of performance. There was no interruption of the practice session with the matching procedure. The groups were roughly equated for sex. Each *S* worked singly in a quiet room.

Apparatus.—A Koerth-type pursuit rotor which turned clockwise at 60 rpm was employed. The black rotor disk was 28.5 cm. in diameter; the brass target was 1.9 cm. in diameter, and was set 8.5 cm. from the center of the larger disk. Two Standard Electric timers recorded time on target in units of .01 sec. alternately every 10 sec. The *E* readily accomplished this by manual operation of a toggle switch, recording time on target and resetting one timer to zero while the other timer was in circuit.

Procedure.—The total practice session was divided into three work periods of 6, 12, and

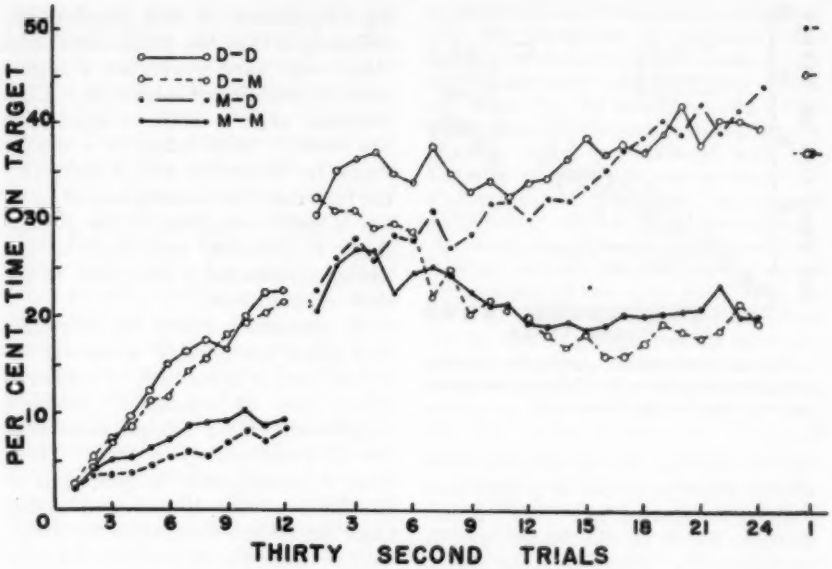


FIG. 1. Performance curves for the first two practice periods and the score on the first trial of the third period.

3 min., in that order, separated by two rest periods of 5 min. and 3 min., respectively. It was necessary to keep the length of the entire experimental session within 50 min. in order to obtain *Ss*. Massed practice (M) consisted of continuous practice while distributed or spaced practice (D) consisted of alternating intervals of 30-sec. work and 30-sec rest. The conditions under which one of each of the eight subgroups practiced during the three work periods were as follows: D–D–D, D–D–M, D–M–D, D–M–M, M–M–M, M–M–D, M–D–M, M–D–D. This gave two different groups during the first practice period, four in the second, and eight in the third.

All *Ss* were given the customary instructions and demonstration by *E* as to the operation of the rotary pursuit device. During spaced practice, *S* stood in front of the rotor during the rest interval and received a signal to pick up the stylus 2 sec. before the beginning of each trial. During the longer rest periods, *S* sat down and conversed or read. Prior to a shift in conditions, *S* was informed of the change. The *S* always began a trial with the rotor in motion.

RESULTS AND DISCUSSION

Time on target was converted into percentage of time on target. For the analysis of variance it was necessary to transform the percentages into angles with mean and variance independent as recommended by Snedecor (9). Reminiscence, or gain over rest, was measured as the difference between the last prerest trial and the first postrest trial.[1] Gain scores were not converted into percentages and an analysis of variance was performed directly. According to Bartlett's test of homogeneity, the *F* test was at all times applicable.

Figures 1 and 2 indicate that the groups were well matched at the beginning of practice ($F = .11$, $df = 7$ and 56) and continued to be well matched under like conditions throughout the first practice period.

[1] This is a common measure of reminiscence which is especially suitable for the present comparisons; although it does not take into consideration the practice effect of the previous trial this is more than compensated for by the failure to include warm-up effect.

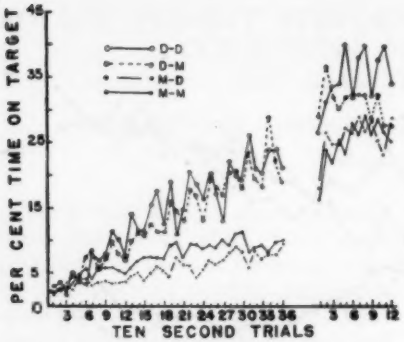


FIG. 2. Performance curves for the first practice period and the first 2 min. of the second period as based on the 10-sec. unit.

As is typical, the D groups were clearly superior to the M groups during this first period ($p = .001$). Reminiscence based on the 30-sec. unit is evident in Fig. 1 after both the 5-min. and 3-min rest intervals. Both sets of reminiscence scores are significant beyond the .01 level for all four groups (t's range from 3.24 to 9.86, $df = 15$).

Inspection of Fig. 1 shows that after approximately 4 min. of practice in the second work period the performance of the D–M group has fallen to the level of the M–M group and after about 6 min. the performance of the M–D group has risen to the level of the D–D group. Also the groups practicing under like conditions remain at the same level of performance; the F ratio for this comparison on the last 30-sec. trial of the second period is .75, $df = 2$ and 60. On the other hand, the difference between the two conditions at the end of this period yields a highly significant F ratio of 76.55, $df = 1$ and 60. Hypothesis 1 is clearly supported. Likewise in Weaver's study (**10**), although neither set of groups had converged by the eighth and final minute of practice in the second period, they were not significantly different. An interesting implication of this possible discrepancy is that the counterclockwise rotor may have generated a higher level of conditioned inhibition. Convergence after a shift in conditions has recently been found in a similar study by Reynolds and Adams (**8**); the fact that the convergence or overlap is more complete in the present study is probably the result of our having introduced a rest prior to the shift in conditions.

As commonly found by investigators using comparable work-rest intervals and a gross unit of measurement, the M condition produces significantly more reminiscence than the D condition ($p = .001$). However, if reminiscence is computed on the basis of the 10-sec. unit rather than the 30-sec. unit, then the difference between the M and D conditions becomes negligible. This is particularly true after the 5-min. rest interval when the D groups are still well below the performance ceiling. Not only is this difference not significant ($F = .33$, $df = 1$ and 60), but what difference there is, is in favor of the D condition. Also it should be noted that all of these 10-sec. reminiscence scores, though relatively free of warm-up effects, are significant at the .01 level. These results are essentially the same as those obtained by Weaver (**10**) and are basic to the derivation of the index of conditioned inhibition.

Derivation of the index of conditioned inhibition.—Findings relevant to the derivation of this index for the first practice period are listed below: (*a*) As indicated above there is no appreciable difference between the 10-sec. reminiscence scores of the M and D conditions. (*b*) Reminiscence occurs after the very first 30-sec. practice trial and the first 30-sec. rest interval in the D condition, i.e., in

terms of the first 10-sec. unit of a 30-sec. trial being significantly higher than the third 10-sec. unit of the preceding trial, and continues to occur after every 30-sec. rest. Concomitantly, there is a steady decline in performance within each 30-sec. trial on successive 10-sec. intervals (the third or last 10-sec. unit is always lower than the first). This periodicity of performance becomes quite regular while no such rhythm is evident in the M condition. The resultant "sawtooth" effect in the curves of the D groups is readily seen in the first practice period in Fig. 2. (*c*) The amplitude of the "sawtooth" curve is a constant, except for random variation, after about the first 2.5 min. of practice (Fig. 2). This is true of Weaver's study (**10**) after about 2 min. of practice and the wave effect is even more pronounced than here. Other evidence that temporary inhibition approaches a maximum value in about 2.5 min. is also in the literature (**2, 4**). (*d*) The difference (10%) between the M and D groups on the first 10-sec. or 30-sec. postrest trial of the second practice period, which is an accepted measure of conditioned inhibition, is equal to the difference (10%) between the M and D groups on the 36th 10-sec. trial of the first practice period (Figs. 1 and 2). (*e*) Lastly, it should be noted that our interpretation is uncomplicated by any difference in the degree of postrest warm up between the M and D conditions, as measured by either the increase in performance within the first 30 sec. of postrest practice or by the increase from the first 30-sec. trial to the peak of the early postrest performance (Figs. 1 and 2).

Three assumptions are basic to our derivation of conditioned inhibition. We assume that the total difference between M and D performance may be attributed to temporary and conditioned inhibition, that the dissipation of temporary inhibition is complete within 5 min., and that conditioned inhibition only builds up under the M condition (**6**). The convergence of the curves in the second practice period is consonant with the first two assumptions, and at least two studies (**2, 10**)[2] support the second assumption. Given these assumptions, the data, in a consistent interrelated fashion, support the inference that on every third 10-sec. trial from Trial 15 to Trial 36 the total difference between the M and D groups consists of conditioned inhibition. The conditions requisite to making this inference and the references to supporting evidence are: (I) The amount of temporary inhibition present at the end of a 30-sec. practice trial should be the same for both conditions (Finding *a*). (II) The amount of conditioned inhibition in the M condition prior to rest should be equal to the amount of conditioned inhibition present after rest (Finding *d*). (III) Conditions I and II imply that temporary inhibition builds up more rapidly than it dissipates in the D group until a maximal level is reached which is the same as the maximal level attained under the M condition, but that once the maximal level has been attained the amount of temporary inhibition dissipated in 30 sec. and the amount built up in 30 sec. are equal (Findings *b* and *c*).

At this point it should be noted that the selection of the 30-sec. work and 30-sec. rest schedule was fortuitous in terms of satisfying conditions I, II, and III. In this connection, it is important to realize that using a 5-sec. work and 25-sec. rest schedule, for example, which would improve performance in the D condition and increase the difference between the M and D condition would *not* produce a greater index of condi-

[2] In a preliminary study, Weaver found that a 5-min. rest yielded insignificantly more reminiscence than a 10-min. rest for a 6-min. prerest practice period. This is the same as Ammons' finding in the closely comparable situation for an 8-min. prerest practice period.

tioned inhibition because the conditions above could not be satisfied.

Thus, to reiterate, we are stating that given all these findings and assumptions it follows that the difference between the lowest point of each 30-sec. trial on the D curve (every third 10-sec. trial after 2.5 min. of prerest practice) and the corresponding point on the M curve is entirely due to the conditioned inhibition built up solely in the M condition. Accordingly, we have plotted the difference between the mean of the D groups and the mean of the M groups on every third 10-sec. trial of the first practice period and drawn a negatively accelerated curve visually through the points, beginning somewhere near the spot that temporary inhibition no longer contributes to the difference. The resultant curve of the development of conditioned inhibition is shown in Fig. 3 where it is obvious that a straight line would fit the points as well. However, the curve is similar to those derived by others (**4, 10**).

Theoretically, if one considers, as we do, that the unconditioned stimulus for the establishment of conditioned inhibition *is* the massing condition, then when massing (US) is omitted by introducing distributed practice the conditioned inhibition, like other conditioned responses, should undergo extinction. Since there is no warm-up differential between the M and D groups at the beginning of the second practice period

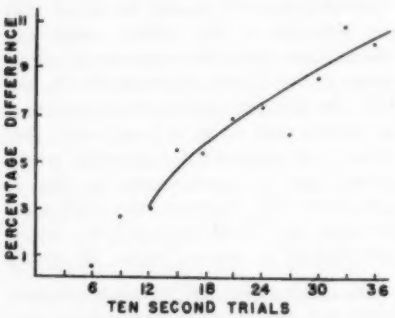


FIG. 3. Derived developmental curve of conditioned inhibition in the first practice period in the M group.

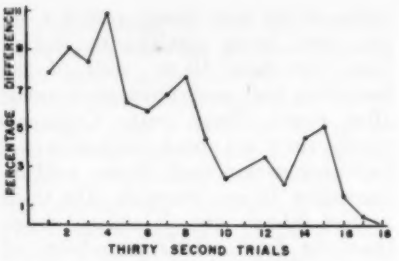


FIG. 4. Extinction of conditioned inhibition in the second practice period in the M–D group.

we can assume and be entirely consistent with the formulation above that the ascent of the M–D group to the level of the D–D group represents the extinction of the conditioned inhibition which was established under M conditions.[3] Therefore, the differences between M–D and D–D on every 30-sec. trial to the point of equality have been plotted in Fig. 4. We see here a rather typical extinction curve. Hypothesis 2 has received convincing support.

Other considerations.—Additional support for the interpretation that the convergence of M–D and D–D represents the extinction of conditioned inhibition in M–D arises from an evaluation of Hypothesis 3, which states that the M–D–M group will show the most immediate decline in performance in the third practice period. The hypothesis has been evaluated graphically in Fig. 5, instead of statistically, as being more meaningful. The most crucial comparison is between M–D–M and D–D–M, the only two groups which were shifted from the D condition to the M condition at this point. The trends of the first

[3] On the basis of the available evidence it does not seem reasonable to assume that the M–D group "catches up" with the D–D group during the second practice period because of the dissipation of temporary inhibition which was not dissipated during the 5-min. rest or because the M–D group is receiving differentially greater increments to the rotary pursuit habit than the D–D group. The remaining extinction of conditioned inhibition hypothesis appears to us to be more parsimonious and tenable.

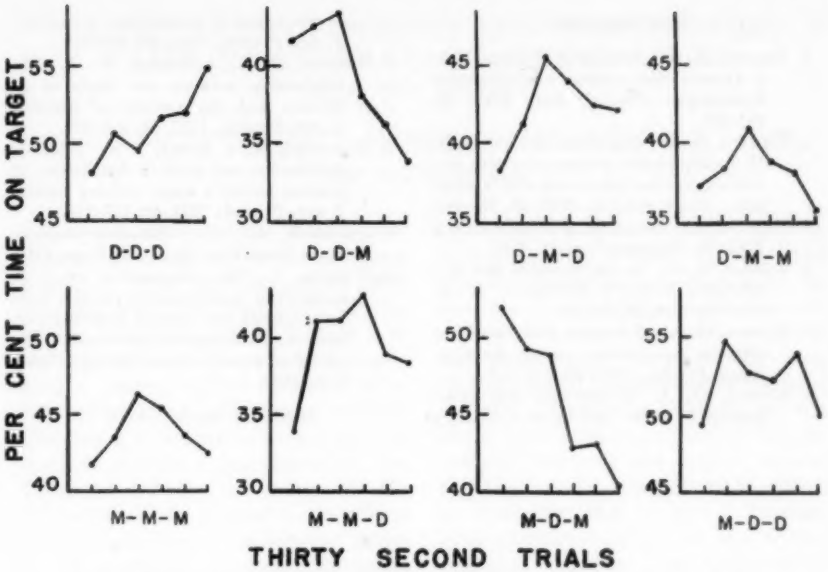


FIG. 5. Performance curves for all eight groups in the third (3 min.) practice period.

three trials for these two groups are in the opposite direction: M–D–M is the only group which does not exhibit warm up. This evidence for rapid reconditioning of inhibition in M–D–M seems to offer indirect support for postulating the extinction of conditioned inhibition during the D condition.

Finally, through the use of the 10-sec. unit we can evaluate the notion of immediate postrest warm up. In Fig. 2 it is apparent that the third 10-sec. unit of the first 30-sec. postrest trial of the second work period is higher than the first 10-sec. unit, whereas in the case of the prerest 30-sec. trials the opposite relationship is true. The t for the difference between the first and third 10-sec. units of the first postrest 30-sec. trial for all groups considered together is 3.58 ($p = .001$). Hypothesis 4 which asserts an immediate warm-up effect has been substantiated.

The authors make no assumptions as to the effector basis of temporary inhibition or as to the way conditioned inhibition is mediated. According to the present analysis, however, it would appear that motor learning phenomena fit into a larger framework of conditioning data and learning theory.

SUMMARY

Sixty-four *Ss* practiced the rotary pursuit task in three different periods separated by rests, with the additional variable of massed (M) vs. distributed (D) practice. This gave a total of eight groups: M–M–M, M–M–D, M–D–M, M–D–D, D–D–D, D–D–M, D–M–D, D–M–M. Massed practice was continuous, and distributed was made up of alternate intervals of 30-sec. work and 30-sec. rest. The rotor turned clockwise at 60 rpm, and performance was measured in 10-sec. units.

Several independent lines of evidence supported the notion of conditioned inhibition in motor learning. It was possible to plot its development during massed conditions and its extinction during subsequent distributed conditions. Temporary inhibition built up rapidly to a maximal level even under distributed conditions, and approximately 50% of this amount dissipated within a 30-sec. rest. Postrest warm up was found to begin immediately.

References

1. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
5. Kimble, G. A. A further analysis of the variables in cyclical motor learning. *J. exp. Psychol.*, 1949, **39**, 332–337.
6. Kimble, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.
7. Kimble, G. A., & Shatel, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, **44**, 355–359.
8. Reynolds, B., & Adams, J. A. Effect of distribution and shift in distribution of practice within a single training session. *J. exp. Psychol.*, 1953, **46**, 137–145.
9. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Iowa State Coll. Press, 1946.
10. Weaver, J. The comparative effect of massed and spaced pre-rest practice upon both massed and spaced post-rest performance on the pursuit rotor task. Unpublished master's thesis, Michigan State Coll., 1950.

(Received March 5, 1954)

Journal of Experimental Psychology
Vol. 49, No. 1, 1955

A COMPARISON OF PURSUIT AND COMPENSATORY TRACKING UNDER CONDITIONS OF AIDING AND NO AIDING

RUBE CHERNIKOFF, HENRY P. BIRMINGHAM, AND FRANKLIN V. TAYLOR

Naval Research Laboratory

Previous studies have shown that pursuit tracking yields lower error scores, or more time on target, than does compensatory tracking. In the pursuit mode of tracking, *S*'s task is to position a control marker to coincide with a course marker which is independently driven. In compensatory tracking *S* attempts to keep a moving marker, which is positioned by the difference between the course and the control outputs, aligned with a fixed reference marker located in the center of the display. Poulton (**4**) found that the mean error in pursuit tracking was about half that in compensatory, and that this difference held up for both a simple and a complex course and through a range of course speeds. Senders and Cruzen (**6**) considered the pure pursuit and pure compensatory conditions as limiting cases of a continuum and compared tasks containing pursuit and compensatory components in various proportions. Their results showed that tracking performance improved continuously and significantly as the amount of the pursuit component in the task was increased from zero to 75%, leveling off thereafter.

The studies referred to above which compared the compensatory and pursuit modes of tracking both utilized a "position," or "unaided" tracking control system in which the output position of the marker is proportional to the input position of the control. The superiority of pursuit tracking over compensatory with this type of system may be accounted for in terms of the information available to *S* in each of the two modes. In pursuit, *S* sees the course as it is generated; he sees the result of his control movements, and he is continually aware of any discrepancy between the two. However, with compensatory tracking the only information *S* can act upon is the displacement error which has developed between the course motion and the control movements he makes to null out this motion. Further, with pursuit tracking *S* can utilize position, velocity, and acceleration changes in the course, the control marker, and the error as cues for predicting the future path of the course and adjust his response motions accordingly. In the compensatory situation, this prediction is made difficult, since the position, velocity, and acceleration changes of the course are confounded with the changes in *S*'s control movements. Thus, in compensatory, only position, velocity, and acceleration changes in the error can be used as cues.

In gunfire control, "aided-tracking" systems are customarily employed since they permit very accurate tracking. With such systems, a movement of the control not only causes a change in the position of the control marker, but also introduces a change in the velocity and, in some systems, the higher derivatives of the controlled quantity. Searle (**5**) has demonstrated that, at least under certain test conditions, the addition of an acceleration term to a combination of position and velocity definitely im-

proves the accuracy of tracking. The ratio expressing the proportional sensitivities of the components of control is usually referred to as the "aiding ratio." With the proper ratio, *S* can correct an error with a control motion proportional to the position component of error—i.e., he can respond to the error as if it were momentarily stationary—and thereby set in changes in velocity and the higher derivatives (if they were included in the system) in amounts which are correct on the average to match those of the course. For fuller discussions of aiding, see Birmingham and Taylor (**1**), Mechler, Russell, and Preston (**3**), and Searle (**5**).

Since with aided tracking *S* need not supply constant velocities or accelerations, it may be hypothesized that the cues which *S* uses to control these aspects of his response will be no longer needed. If this should be so, it would be expected that (*a*) no difference in accuracy will be found between pursuit and compensatory tracking if proper aiding is employed in the tracking system, and (*b*) aiding will have a beneficial effect upon both tracking modes. The present experiment was designed to test these hypotheses.

Method

Apparatus

Display.—A 5-in. cathode ray tube (CRT), with a P-11 short persistence screen, was used to display the tracking information. By means of an electronic switch, the CRT beam was time-shared to produce two spots of light. One spot was shaped into a $\frac{1}{4}$-in. vertical line, while the other was left as a $\frac{1}{16}$-in. diameter dot. During pursuit tracking, the vertical line was controlled by *S* and the dot positioned by a course generator. When compensatory tracking was used, the dot was fixed in the center of the CRT face as the reference, while the vertical line was driven by the difference between the course generator output and *S*'s control movements.

Control.—The vertical line marker was positioned by *S* by means of a handwheel mounted on a helipot variable resistor. The handwheel was a 6-in.–diameter aluminum disk with a $\frac{3}{4}$-in.–diameter knob fastened at the rim. With the unaided tracking system the handwheel-to-marker sensitivity was set so that when the handwheel was rotated through 100°, the marker moved a distance of 1 in. across the face of the CRT. With aided tracking, the relationship between the handwheel and marker was such that a 100° rotation of the handwheel produced a position change of 1 in., a velocity change of 2 in./sec., and an acceleration change of 3 in./sec.^{2} Thus, in this study the aiding ratio used was 1:2:3.

Course.—The course motion was a complex of three sine waves, of 6, 10, and 15 cycles per min. The relative amplitude values for the three sine waves were 20, 10, and 1, respectively. The maximum course excursion was across the center 4 in. of the CRT face.

Scoring.—Since an AC voltage, which is proportional in amplitude to the separation between the spots on the CRT face, exists at the output of the electronic switch, it was possible to obtain directly a measure of spot separation. By use of an electronic integrator, this voltage was integrated over the scoring duration by cumulating a charge on a condenser. Integrated error was then read on a microammeter.

Procedure

The *S* sat facing the CRT display at a viewing distance of approximately 12 in. A fluorescent desk lamp was placed behind the CRT and facing the wall to provide room illumination with a minimum of reflection from the CRT face. The handwheel was mounted 10 in. to the right of the CRT centerline, and at the seated *S*'s elbow height.

No attempt was made to train *S* to respond in any specific manner to the various conditions. Before the experiment proper was begun, one training session with seven practice trials per condition was given to familiarize *S* with the equipment. For pursuit tracking, *S* was instructed to turn the handwheel so as to keep the vertical line marker on the moving course dot, while for compensatory he was instructed to keep the vertical line marker on the fixed reference dot.

Six Navy enlisted men served as *S*s. The four conditions, compensatory unaided (CU), compensatory aided (CA), pursuit unaided (PU), and pursuit aided (PA), were presented in randomized order in blocks of six trials per condition for ten daily sessions. One practice trial always preceded each block of scored trials. A trial was 1 min. in duration, with the last 45 sec. scored by the error integrator. The *S* was given a 30-sec. rest interval between trials, while the

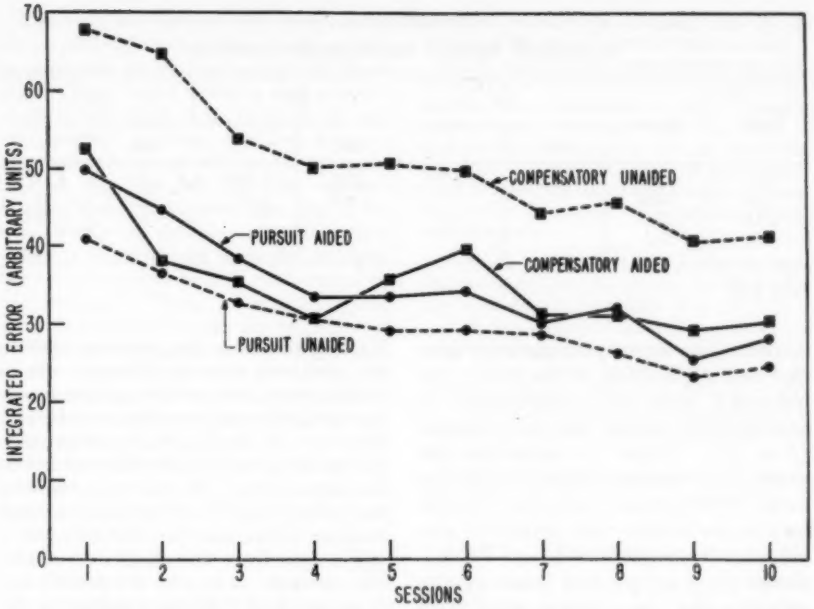


FIG. 1. Tracking error scores for the four experimental conditions during the ten sessions. Each point represents the mean of six trials for each of six *S*s.

interval between successive conditions was approximately 1½ hr.

The 1-min. course duration and 45-sec. scoring interval were automatically programmed. After each trial, *E* recorded the integrated-error score cleared the meter by momentarily shorting the integrating condenser, and waited 30 sec. before starting the next trial. Also before each trial, *E* positioned the course dot near the center of the CRT face and randomly varied the direction in which the course would start.

RESULTS

Tracking effectiveness with each of the tracking systems was measured in terms of integrated-error scores. This method of scoring provides a number expressing the total cumulative error integrated over time. The measuring units, while arbitrary, are on a linear scale. Figure 1 presents the graph showing integrated error scores for the four experimental conditions for each of the ten daily sessions. Since the main interest in the study was comparative tracking performance after *S* had reached a fairly stable level, all statistical analyses have been performed on the data taken during the last four days of the experiment (Sessions 7–10).

Inspection of the integrated-error tracking scores revealed that their

TABLE 1

ANALYSIS OF VARIANCE OF TRACKING ERROR SCORES

Source of Variation	*df*	Mean Square	*F*
Subjects (Su)	5	785.03	78.27**
Conditions (Co)	3	1334.26	133.03**
Sessions (Se)	3	94.61	9.43**
Su × Co	15	54.66	5.45**
Su × Se	15	13.38	1.33
Co × Se	9	8.02	.80
Su × Co × Se	45	10.03	—

** $p < .01$.

TABLE 2

COMPARISONS BETWEEN EXPERIMENTAL CONDITIONS

Cond.	Means	CA		PU		PA	
		Diff.	t	Diff.	t	Diff.	t
CU	42.8	12.4	5.44**	17.0	5.54**	13.7	4.87**
CA	30.4			4.6	2.85*	1.3	.93
PU	25.8					3.3	6.08**
PA	29.1						

* $p < .05$.
** $p < .01$.

distributions were approximately normal and application of Bartlett's test indicated that the assumption of homogeneity could not be rejected ($p > .20$). Table 1 summarizes the analysis of variance performed on the data. Differences between experimental conditions were tested for significance by use of t tests, and Table 2 shows the t values and levels of significance for the various conditions compared.

From the curves in Fig. 1 and the t values of Table 2, it can be seen that the compensatory unaided-tracking system produced more error than any of the other systems. On the other hand, the pursuit unaided system produced the least amount of tracking error. Comparison of the pursuit aided- and the compensatory aided-tracking modes showed no significant difference in error scores. Applying aiding thus resulted in a more accurate compensatory tracking system, but a less accurate pursuit system. This differential effect of aiding on the two tracking modes acted to bring them together to the same accuracy level.

DISCUSSION

The results of the present study corroborate the findings of Poulton (4) and of Senders and Cruzen (6) in that with position control (no aiding), pursuit was clearly superior to compensatory tracking. At the same time, the data confirm the prediction that no difference will be found between the pursuit and compensatory modes if aided tracking is employed. However, the finding that unaided pursuit tracking excelled all other conditions was unexpected. So also was the fact that aiding caused a worsening in pursuit tracking rather than an improvement.

The data from this study do not provide evidence as to why the pursuit unaided mode of tracking resulted in the least amount of error. However, a possible explanation would be that the aiding ratio used in this study was not such as to bring about optimum improvement with aiding. Searle (5) studied various combinations of component sensitivities and found an aiding ratio of 1:4:8 to be best for two experiments and an aiding ratio of 1:2:8 to be best for the third. A misadjustment of the aiding parameters in the present study resulted in an aiding ratio of 1:2:3 instead of the intended 1:4:8. The possibility that a better aiding ratio may bring the compensatory aided scores down to, or below, that of pursuit unaided, will be investigated.

The actual worsening of pursuit tracking when aiding was employed may be due to *S*'s inability to restrain himself from responding to the velocity information which is now relatively uncorrelated with tracking error. Thus, the very factor making for the superiority of the pursuit display in the unaided-tracking situation may act as a distraction in the presence of aiding to cause a deterioration of performance.

Over-all, the results lend credence to the hypothesis that pursuit tracking will be superior to compensatory in those cases where target velocity and acceleration information is required by *S*, but that pursuit tracking will lose its pronounced advantage when this information is not needed by the tracker. Furthermore, it appears that the inclusion within a display of information which is uncorrelated with the tracking behavior required of *S* is detrimental to performance. Thus, it would appear that recommendations as to whether pursuit or compensatory tracking should be used in specific systems must take into account the precise nature of the informational requirements of the particular tracking task involved.

The finding that aided pursuit tracking is inferior to unaided pursuit tracking supports the results of a study by Lincoln and Smith (2). In comparing position, velocity, and aided tracking with a pursuit display they found that position tracking was superior at all stages of practice to either velocity or aided tracking. Their conclusion that aiding hampered rather than enhanced tracking was unique at the time since engineers had been designing tracking systems on the basis of tests which showed that aiding, in fact, did aid. However, all of this previous engineering evidence had been collected on compensatory tracking systems, whereas the study by Lincoln and Smith was carried out with pursuit tracking. The results of the present study now make clear that the apparent contradiction is resolved if the mode in which the tracking information is presented to *S* is recognized as an important dimension of tracking behavior.

Summary

Previous studies have shown that pursuit tracking yields lower error scores than compensatory. This experiment was designed to test the hypotheses that aiding applied to compensatory and pursuit tracking will result in (*a*) improved performance on both, and (*b*) no difference in accuracy between the two modes.

Six *S*s were given six trials daily for ten days on each of the four conditions: (*a*) compensatory unaided, (*b*) compensatory aided, (*c*) pursuit unaided, and (*d*) pursuit aided tracking. A trial was 1 min. in duration, with the last 45 sec. scored by an error integrator. A course composed of three sine waves was tracked by means of a handwheel control. The following results were obtained:

1. As in previous studies, pursuit tracking resulted in significantly less error than compensatory when no aiding was used ($p < .01$).
2. Aiding materially reduced the amount of error in the compensatory situation ($p < .01$).
3. Aiding applied to the pursuit mode increased the amount of tracking error ($p < .01$).
4. With aiding, there was no significant difference in error between pursuit and compensatory tracking.
5. Pursuit unaided tracking was significantly better than any of the other three conditions.

References

1. Birmingham, H. P., & Taylor, F. V. A human engineering approach to the design of man-operated continuous control systems. *U. S. Naval Res. Lab. Rep.*, 1954, No. 4333.
2. Lincoln, R. S., & Smith, K. U. Systematic analysis of factors determining accuracy in visual tracking. *Science*, 1952, **116** (3008), 183–187.
3. Mechler, E. A., Russell, J. B., & Preston, M. G. The basis for the optimum aided-tracking time constant. *J. Franklin Inst.*, 1949, **248**, 327–334.
4. Poulton, E. C. Perceptual anticipation in tracking with two-pointer and one-pointer displays. *Brit. J. Psychol.*, 1952, **43**, 222–229.
5. Searle, L. V. Psychological studies of tracking behavior. Part IV—The intermittency hypothesis as a basis for predicting optimum aided-tracking time constants. *U. S. Naval Res. Lab. Rep.*, 1954, No. 4333.
6. Senders, J. W., & Cruzen, M. Tracking performance on combined compensatory and pursuit tasks. *USAF, WADC Tech. Rep.*, 1952, No. 52-39.

(Received April 12, 1954)

Journal of Experimental Psychology
Vol. 49, No. 1, 1955

MITIGATION OF WORK DECREMENT

G. T. HAUTY AND R. B. PAYNE

USAF School of Aviation Medicine

The compensatory pursuit task provided by the USAF SAM Multidimensional Pursuit Test is particularly suited to the study of work decrement and its counteraction. Inherent in this task are complex perceptual and motor requirements resembling those of skilled work. In addition to practical relevancy, the task yields an objective and precise index of work decrement. Early performance reveals rapid and significant gains in proficiency, but continued practice results in a gradual rate of decrement despite the profession of sustained interest in the task.

The present investigation was designed primarily to reveal whether or not such work decrement could be modified by certain psychological and physiological conditions. The former consisted of different instructions concerning the duration of the work period and different methods of presenting information signifying performance adequacy. The physiological conditions, which consisted of cerebral analepsis and depression induced by oral consumption of appropriate drugs, have been investigated earlier (**5, 6**). The former treatment significantly allayed the usual decline in proficiency and the latter significantly hastened and enhanced it. Remaining to be studied, however, were the specific questions concerning work output following the dissipation of these effects, the possibilities suggested by properly timed, repetitive analeptic dosages, and the pharmacological counteraction of depressant effects. A secondary objective concerned the transferability of effects induced during the prolonged task to performance on a dissimilar perceptual-motor task.

METHOD

Apparatus.—The USAF SAM Multidimensional Pursuit Test (**4**) requires *S* to correct for the random drift of four simulated aircraft instruments from their null positions by the timely and appropriate manipulation of four simulated aircraft controls. A timer is activated only when all four instruments are concurrently maintained in their null positions. A cycling system meters out alternate work trials (1 min.) and rest periods (15 sec.) for any desired span of time. By modification of the apparatus, supplementary cues signifying concurrent maintenance of the four instruments at null could be presented to *S*. These were a visual signal, given by an unobtrusively positioned lamp which illuminated the floor of the cockpit, or an auditory signal consisting of a 400-cycle tone transmitted through a very light-weight monaural receiver.

The USAF SAM Self-Pacing Discrimination Reaction Time Test (**3**) was used for the terminal and dissimilar perceptual-motor task. Essentially, *S* is presented with one of four spatially different pairs of red and green lights to which he must respond by quickly throwing one of four switches on the basis of the position of the green light relative to the red light. A cycling system randomly presents the stimuli for 2-min. trials separated by 30-sec. rest periods. Correct and incorrect responses are electrically accumulated.

Experimental conditions.—The three variables to be investigated were of the following nature:

a. Pharmacological treatments, orally administered in capsule form, which involved the subclasses: D_1—a control group receiving no drug; D_2—placebo (lactose); D_3—Dexedrine, single dose (5 mg. d-amphetamine sulfate); D_4—Dexedrine, repetitive dose (timed disintegration with 4-hr. peak release of approximately 6½ mg. d-amphetamine sulfate per disintegration); D_5—caffeine derivative; D_6—Benadryl-Hyoscine-Dexedrine (50 mg. diphenhydramine hydrochloride, .65 mg. scopolamine, 5 mg. d-amphetamine sulfate); D_7—Benadryl-Hyoscine

(50 mg. diphenhydramine hydrochloride, .65 mg. scopolamine).

b. Different systems of information as to when all four instruments were being concurrently maintained in null positions. This was to be primarily given by I_1—direct observation of the positions of each of the instruments; I_2—a visual signal; or I_3—an auditory signal.

c. Different goal proximities as determined by knowledge of the length of work task. Before the start of the work task, half of the *Ss* were informed, G_1—"You will be required to perform this task for a period of 4 hr. after which you will be given a 'break.'" The other half of the *Ss* were informed, G_2—"You will be required to perform this task for a period of 7 hr."

Procedure.—All *Ss* received, under common conditions, 40 trials of initial training in order to get beyond the early learning increments. Immediately following this, *Ss* were given the instructions and drugs specific to the particular experimental conditions and they then began the initial period of work which consisted of 192 trials (9:00 A.M. to 1:30 P.M.). At the completion of Trial 192, all *Ss* received a 15-min. rest during which they left the testing rooms, relieved themselves, and ate a light lunch consisting of a sandwich, a candy bar, and ½ pt. of milk. They then returned to the apparatus formerly occupied and began the final work period consisting of 136 trials (1:45 to 4:45 P.M.). In both the initial learning and work periods, *Ss*' performance for each eight trials of work (1 cycle) was recorded. At the conclusion of the prolonged task, *Ss* received operational instructions for the Discrimination Reaction Time Test and were then given eight 2-min. trials on this apparatus (4:50 to 5:10 P.M.). Correct and incorrect responses were recorded after each trial.

The *Ss*, randomly assigned to the 42 cells ($N = 4$) yielded by the factorial design, were tested four at a time under conditions which prevented observation or knowledge of one another's performance, and they were treated alike for any one testing day. The cells were randomized with respect to the serial order of testing days. Illumination, temperature, and humidity were maintained at conditions considered optimal.

Subjects.—The 168 *Ss* were male basic trainees at Lackland Air Force Base. Those designated to receive a drug were given an opportunity to refuse participation in the investigation. Only one *S* did refuse thus necessitating the testing of a replacement. It was possible to supervise *Ss*' activities for a period of 22 hr. prior to their participation. Consequently, it is reasonable to assume that *Ss* began the investigation in a normal physical state.

Results

Statistical analyses.—The individual measures of proficiency obtained during every third cycle of the 4-hr. prerest and 3-hr. postrest work periods were adjusted for regression upon the mean proficiency achieved during the last three pretreatment cycles. Heterogeneity attributable to a progressively inverse relationship between subclass means and standard deviations was satisfactorily reduced by the transformation $y = 2 \arcsin \sqrt{x}$. The variation of the transformed scores was then analyzed in accordance with the split-plot design of the investigation, where Goal Proximity (2) × Information Systems (3) × Drugs (7) constitute the independent experimental variables and Cycles (16) are treated as replications. The results

TABLE 1

Analysis of the Adjusted Variation of Transformed Proficiency Measures

Source	*df*	Mean Square	*F*
Goals	1	5.434	18.74***
Information	2	7.639	26.34***
Drugs	6	7.260	10.58***
G × I	2	.710	2.45
G × D	6	.686	2.37*
I × D	12	.280	0.97
G × I × D	12	.403	1.39
Ss treated alike	126	.290	
Total between *Ss*	167		
Cycles	15	2.644	132.20***
C × G	15	.028	1.40
C × I	30	.033	1.65*
C × D	90	.125	3.79**
C × G × I	30	.012	0.60
C × G × D	90	.033	1.65**
C × I × D	180	.014	0.70
C × G × I × D	180	.020	1.67**
Pooled error	1890	.012	
Total within *Ss*	2520		
Total	2687		

* Significant at .05 level of confidence.
** Significant at .01 level of confidence.
*** Significant at .001 level of confidence.

are presented in Table 1. Since only one of the variables, Goal Proximity, was considered to be a random variable, the main effects, G and I, were tested by Ss treated alike while main effect D was tested by the interaction G × D. The C × G × I × D interaction was used to test the triple interactions, the simple interactions, C × G and C × I, and the main effect C. The C × D interaction was tested by the interaction C × G × D.

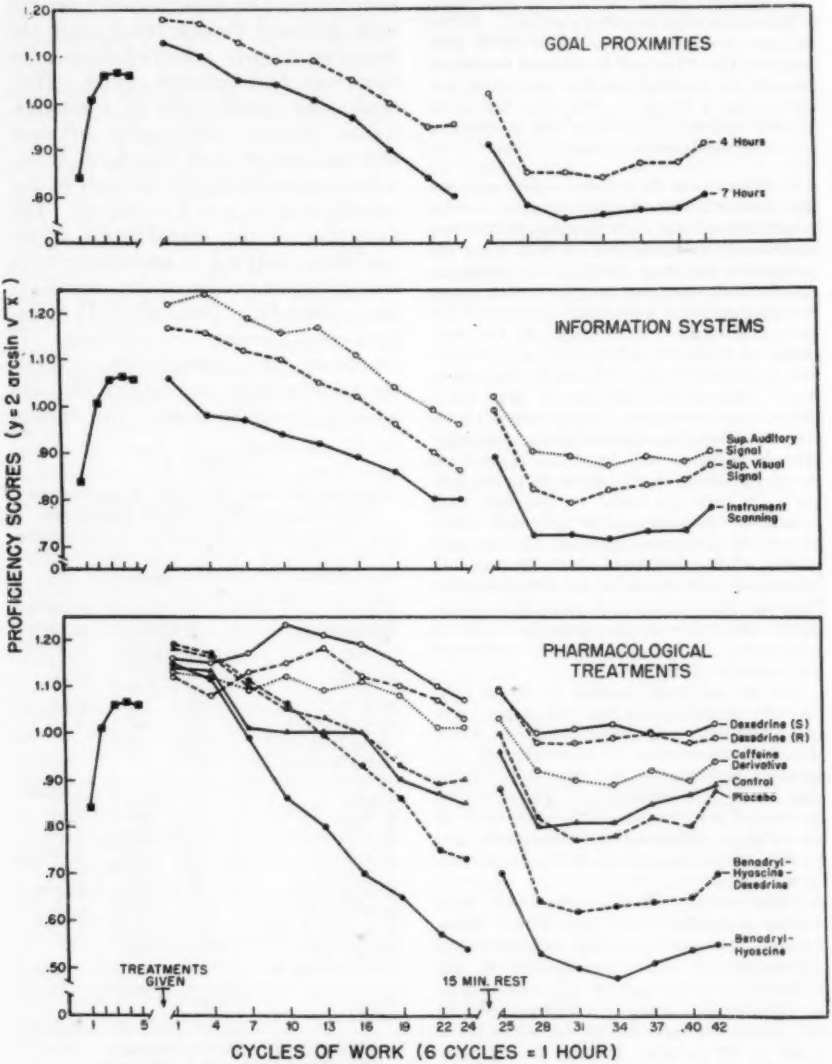


FIG. 1. The effects of the respective conditions of each variable upon work proficiency.

Next, for each of the three variables the means of their respective conditions were determined for each of the 16 cycles analyzed. These were plotted to give the work curves (see Fig. 1), which indicate clearly the magnitude of experimentally induced differences and the interaction of these differences with the duration of work. Finally, to obtain a more definitive interpretation of this interaction, detailed tests of significance were computed at certain critical points of the total work period: Cycle 24, which immediately precedes the rest period; Cycle 25, which immediately follows this period; the mean of the performances attained during Cycles 31–34 inclusive, which exhibit the lowest over-all output; and the last cycle of work, 42. The tabular data of each of these four analyzed periods of work, cell and subclass obtained means, subclass means of the transformed adjusted values, analyses of the adjusted variances, and matrices of t tests are given in a more detailed report of this investigation (**2**).

Goal proximity.—The pertinent curves reveal for the first cycle of work a direct relationship between pace of work and proximity of work termination which is sustained up to the end of the prerest 4-hr. period. At this point, it would appear that the difference in proficiency between the two groups ($p < .01$) has been augmented by an increment attributable to a goal gradient effect in the case of the group anticipating the period of rest. Since the more immediate goal was expected to effect a faster pace of work, the question of greater importance was whether or not this superiority would be maintained following the rest period which both groups actually received. The curves, continuing to graphically represent that which is to be noted in Table 1, clearly resolve the question. At the three discrete analyzed periods of postrest work, the groups remain significantly different ($p \leqq .01$). Both groups knew, following the rest period, how much longer they would be required to work; hence, the sustained difference in proficiency must be due to a remarkable persistence of the work pace set by each group. The interaction of this variable with Drugs will be discussed later.

Information systems.—The introduction of a single supplementary signal, auditory or visual, is seen in Fig. 1 to produce initially a marked superiority of performance. As work continues during the first 4 hr., there appears to occur a progressive diminution of superiority which accounts, in large measure, for the significance of the C × I interaction. The instrument-scanning group, however, remains significantly inferior ($p \leqq .05$) at the end of this 4-hr. period, Cycle 24, and during Cycles 25, 31–34, and 42 ($p \leqq .01$). Since the auditory signal had the effect of reducing the visual load, it was expected to lead to performance superior to that resulting from the visual signal. Such is suggested by the consistent ordering of the curves and is demonstrated as a statistically significant fact, the difference between the over-all means for the 16 cycles being significant at better than the .01 level of confidence.

Pharmacological treatments.—The significance of the C × D interaction was expected on the grounds of absorption latencies, differential dissipations, and the antagonistic effects of cerebral analepsis and depression. Accordingly, it was felt that the additional analyses of discrete and critical periods of work would be of particular value in the appraisal of such temporally dependent effects.

Considering first the curves of the control and placebo groups, progressive decrement is seen to occur for the first 4 hr. Following the temporary beneficial effects of the rest period, decrement continues until, after approximately 6 hr. of work, the trend is reversed and the characteristic end spurt is exhibited. In none of the *t* tests for Cycles 24, 25, 31–34, and 42 were the differences between the control and placebo groups statistically significant.

The normally occurring work decrement is seen to be markedly arrested by the analeptic properties of Dexedrine and caffeine. During Cycle 24, the *t* tests revealed each of the analeptic groups to be significantly superior to the control groups ($p < .05$). While Dexedrine, as compared to the caffeine derivative, appears at this point to lead to superior performance, the three analeptic groups are not significantly different.

As revealed in Fig. 1, the period of lowest over-all work output was during Cycles 31–34. Since performances during this period are least likely to be affected by the effects of just having had a rest and by the anticipation of work termination, a detailed analysis of performances should constitute a critical test of the pharmacologically induced differences existing at this point of working time. This period, moreover, being of considerable duration (32 min. of work) gives assurance of reliability. The *t* tests permit the following statements: (*a*) both groups receiving single and repetitive dosages of Dexedrine are significantly superior to the control groups ($p \leqq .01$), (*b*) in comparison, the disproportionately greater decrement demonstrated by the group receiving the caffeine derivative is reflected in the tests of significance which indicate this group to be significantly inferior ($p = .04$) to the Dexedrine (single dosage) group and significantly superior ($p = .03$) to only one of the control groups, the placebo group, and (*c*) no significant difference exists between the groups receiving the single and the repetitive dosages of Dexedrine.

From this point, the output of the groups receiving Dexedrine continues at a relatively constant rate up to the last cycle of work. Here, these two groups remain significantly superior to the control groups whereas the group receiving the caffeine derivative is no longer significantly different from either control group.

The sedative property of Benadryl-Hyoscine is seen to effect a profound rate of decline. For each period analyzed, the group receiving Benadryl-Hyoscine was significantly inferior to the control groups ($p < .01$). The inclusion of Dexedrine in the Benadryl-Hyoscine preparation does have significant effects. The group receiving this treatment, while significantly inferior to the control groups at the points analyzed, was nevertheless significantly superior to the Benadryl-Hyoscine group. For both groups, the rest period is followed by a temporary but marked increment in output. This and the demonstrated end spurts give indication of the fact that trend in work decrement engendered by time at the task and the sedative action of these particular motion-sickness preventives is not irreversible.

It will be noted in Table 1 that not only is the G × D interaction to be regarded as significant but the interaction C × G × D as well. From inspection of the appropriate data, it could readily be concluded that large and significant differences attributable to Goal Proximity occurred under the conditions of Benadryl-Hyoscine-

Dexedrine, Dexedrine (repetitive dosages), and placebo whereas smaller differences, found not to be statistically significant, occurred under the remaining pharmacological conditions. Moreover, this disproportion of G × D differences existed only within two distinct periods, the first hour and the final 2 hr. of work. For the greater period of work, then, the effects attributable to these two variables may be considered to be additive.

Transfer of decrement.—Following the 7-hr. work task, *S*s were required to work for a period of 16 min. on a dissimilar perceptual-motor task under conditions common to all. The mean correct responses for the total number of trials were analyzed by the method of covariance using the Armed Forces Qualification Test score as the independent variable. The *F* ratios for each of the three variables and interactions could not be considered as being significant. In the case of the incorrect responses, homogenization of variances was met by the transformation $y = \sqrt{x}$ and these transformed scores, adjusted for regression upon the Armed Forces Qualification Test score, were analyzed. Again, the *F* ratios for the principal variables and their simple interactions did not approach a level of significance. The triple interaction, however, was significant at the .05 level of confidence.

Discussion

The consistently superior performance attributable to the information systems employing a single, supplementary cue to signify performance adequacy does not appear to be the primary result of an extrinsic factor, e.g., augmentation of motivation elicited by the novelty of the mechanical modification. The sustentation of superiority up to the end of 7 hr. of work suggests the cause to exist within the task itself. While the information at hand does not permit exact and specific determination of how perceptual behavior is altered by the inclusion of such a supplementary signal, perhaps the following explanation may be considered.

In the absence of a single supplementary cue, each of the four spatially separated indicators must be attended to so as to obtain information of performance adequacy. Perceptually, then, the task is unstructured in the sense that no one particular method of scanning is prescribed or followed. Moreover, any scanning method used in this situation is inefficient because the departure of one indicator beyond tolerable limits could remain unnoticed until *S*'s method of scanning returned attention to that indicator. On the other hand, if successful performance is indicated by a single cue which is heard or peripherally perceived, the perceptual task is simplified in that attention is driven in a more systematic manner. The probable method of the now structured task is the fixation of a smaller field, e.g., the two center dials, while relying upon cessation of the supplementary cue which signifies the occurrence of error at the marginal instruments.

The findings relating to the pharmacological treatments are in agreement with those of a previous study (**5**). There are, in addition, certain results which should be noted. The nearly constant rate of work performed during the final 3 hr. by the group receiving the single dosage of Dexedrine indicates either that performance is highly resistant to a subjective state of "letdown" which might be supposed to occur with the dissipation of analeptic effects or that such a state did not occur within the period observed. Secondly, the addition of Dexedrine to the Benadryl-Hyoscine preparation, a motion-sickness preventive, was for the purpose of counteracting depressant side effects. While counteraction was significantly demonstrated, complete nullification was not realized. Lastly, it was assumed, in the case of the multistage Dexedrine capsule,

that the effects of the second liberated dosage would be evidenced during the last 3 hr. of work by greater output as compared to that of the group receiving the single dosage of Dexedrine. Since such was not to be noted, the efficacy of the multistage capsule used in this investigation was not realized. Possibly the first dosage produced an upper limit of effect which, if sustained throughout the latter period of work, would preclude any further increment of output. Before the effectiveness of multistage analeptics can be fully appraised, however, investigation is required of other task situations, alterations of the dosages, and of longer periods of work.

The nontransfer of experimental effects that clearly existed at the very end of the prolonged task to the terminal task raised the question concerning the degree of over-all fatigue engendered by the prolonged period of work. To attempt an answer, an additional comparison was made. In a previous investigation (1), a group of *S*s ($N = 16$), drawn from the same general population, performed the transfer task under conditions comparable to those of the present investigation save the following exceptions: Performance on the task was not preceded by work on any other task and none of the presently concerned experimental treatments had been administered. These *S*s, therefore, may be considered as representing an "unfatigued" group. Their mean correct and incorrect responses, 77.17 and 6.67 respectively, are offered for comparison with the mean correct and incorrect responses, 79.44 and 6.66 respectively, of the "fatigued" *S*s ($N = 168$) of the present investigation. Clearly then, effects which might be attributed to fatigue are not demonstrated in terminal task performance. This finding may be due to either the possibility that physical impairment had not generalized sufficiently to include the neuromuscular components demanded by the dissimilar perceptual-motor task or the possibility that dissimilarity of the task evoked a motivational level which effected an increased expenditure of energy reserves. Perhaps both possibilities are required for full explanation of the phenomena. To be kept in mind is the fact that the terminal task was of a short duration. Had it been continued for a longer period, there may well have been demonstrated a comparable ordering of the divergent output levels witnessed in the prolonged task and an inferior performance by the fatigued as compared to the unfatigued group.

A concluding statement, permitted by the factorial design of the investigation, concerns the joint effectiveness of the methods studied for the mitigation of work decrement. The results demonstrate that as optimal conditions of varied natures are incorporated into the total work situation, work decrement is correspondingly reduced.

Summary

Preliminary training was given to 168 male basic airmen volunteers, whose ages ranged narrowly around 20 yr., on a compensatory pursuit task involving simulated aircraft instruments and controls. They then performed the task for 7 hr. after having been randomly assigned to 42 combinations of 7 pharmacological conditions, 3 systems of presenting information concerning performance adequacy, and 2 differing proximities of goal. The singular and joint effects of these conditions upon task proficiency were appraised statistically at critical points throughout the 7 hr. of work. At the conclusion of this period, *S*s performed a dissimilar perceptual-motor task for 16 min. Transferability of work impairment was appraised.

1. *D-amphetamine* (5 mg.) effectively mitigated work decrement for the entire period of 7 hr. and was as effective for this period as a multistage capsule which released equivalent dosages at 4-hr. intervals. A caffeine derivative was also effective but for a shorter period of time. The addition of d-amphetamine (5 mg.) to a preparation of diphenhydramine hydrochloride (50 mg.) and scopolamine (.65 mg.) had the desired effect of reducing but not eliminating the effects of sedation upon performance.

2. The greater efficiency of perceptual behavior afforded by supplementary signals resulted in a superiority of work proficiency which was sustained for the 7 hr. of work.

3. The pace of work effected by the more immediate goal resulted in superior work proficiency throughout the 7-hr. period.

4. Effects induced during the prolonged task were not transferred to performance on a dissimilar, terminal task.

References

1. Hauty, G. T. Psychological adaptability: standardization of a psychomotor test and consideration of transfer effects. *USAF Sch. Aviat. Med., Proj. Rep.*, 1953, Proj. No. 21-0202-0005 (Rep. No. 2).
2. Hauty, G. T., & Payne, R. B. Methods for the mitigation of work decrement. *USAF Sch. Aviat. Med., Proj. Rep.*, 1953, Proj. No. 21-1601-0004 (Rep. No. 4).
3. Melton, A. W. (Ed.) Apparatus tests. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 4.)
4. Melton, A. W. (Ed.) Apparatus tests. Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 4. Restricted Supplement.)
5. Payne, R. B., & Hauty, G. T. The effects of experimentally induced attitudes upon task proficiency. *J. exp. Psychol.*, 1954, **47**, 267–273.
6. Payne, R. B., & Hauty, G. T. The effects of motion-sickness preventives upon certain perceptual-motor components of the pilot's task. *USAF Sch. Aviat. Med. Proj. Rep.*, 1953, Proj. No. 21-1601-0004 (Rep. No. 3).

(Received March 26, 1954)

Journal of Experimental Psychology
Vol. 49, No. 1, 1955

VARIABLES AFFECTING THE ANGULAR DISPLACEMENT THRESHOLD OF SIMULATED AUDITORY MOVEMENT[1]

R. C. WILCOTT[2]

University of Southern California and U. S. Navy Electronics Laboratory

In 1901 Pierce (**3**) determined the least perceptible movement of a "buzzing sound" with the movement starting from various positions. Using a sound cage and with the velocity of movement supposedly held constant, he obtained the following results: The least distance of movement perceptible ranged from 2.5° with movement from the median plane to 8.7° at the lateral plane. The purpose of the present experiment was to investigate the perception of movement simulated with a binaural listening system. It has been shown that by means of a linear step-wise change in interaural time delay, the definite illusion of auditory movement can be obtained (**2**). This study was designed to determine the minimum distance (displacement from point of origin in terms of the amount of change in interaural delay) under various conditions required to elicit the perception of movement. The variables investigated were: (*a*) frequency components of the signal, (*b*) point of origin of movement in the phenomenal field, and (*c*) velocity of movement.

[1] This report covers work on the joint University of Southern California–U. S. Navy Electronics Laboratory Project No. 2. It is part of a series of investigations under Contract No. NObsr-57435 between the Bureau of Ships, Navy Department, and the University of Southern California in cooperation with the U. S. Navy Electronics Laboratory. The experimental work was carried on at the U. S. Navy Electronics Laboratory.

[2] Now at the Institute for Juvenile Research, Chicago, Illinois.

Procedure

The signal was produced by a thermal noise generator with an attenuator and filter for controlling the intensity and width of the noise band. The turning on and off of the signal and its duration were controlled by a cuer which operated an electronic switch. Beyond the switch the circuit divided into two channels. The linear step-wise change in interaural time delay was produced in one of the signal channels by a 100-step rotary switch coupled to a 100-unit variable electrical delay line. Each unit of the delay line consisted of a carefully constructed low-pass filter which produced 9.0-microseconds delay. The rotary switch was turned by a constant speed motor operating through a variable speed changer.

To determine the effects of various signal frequency components on the movement threshold two noise bands were selected: (*a*) a band ranging from 106 to 1200 cps, and (*b*) a band ranging from 2400 to 3400 cps. The loudness of each of these noise bands was set at approximately 60 db SL. At this intensity each of these noise bands elicited a clear perception of movement.

The effects of movement from two different points of origin in the phenomenal field were also determined as follows: First, a movement which began at the median plane and progressed to the left toward the lateral plane. This was accomplished by sweeping from a point of zero interaural delay toward the point of maximum delay. Second, a movement was used which began at the right lateral position and progressed toward the median plane. This was accomplished by first introducing an interaural delay sufficient to lateralize the signal (90° shift from the median plane) and then sweeping from this point toward the point of zero time delay. The exact amount of interaural delay necessary to lateralize a signal is difficult to determine as subjective judgments of *O*s are usually unreliable. However, a preliminary investigation with the two noise bands used in the experiment indicates that the minimum interaural delay for complete lateralization varied from about 780 to 800 microseconds. On this basis 792 microseconds of interaural delay was used as the point of

origin for the movement from the lateral position.

The third variable was the velocity of movement. Six velocities of movement were selected ranging from a perceptually slow to fast movement. The following velocities (ds/dt), expressed as a change in the amount of interaural delay per unit of time, were chosen: the ds's being 24, 45, 90, 150, 225, and 450 microseconds of change in interaural delay, where dt equals 1 sec. In Fig. 1 these velocities are also expressed in terms of the approximate perceived change in azimuth per second corresponding to each change in interaural delay per second.

Displacement thresholds were obtained under each of the various conditions of the experiment by setting up a constant velocity of movement and varying the duration of the signal. For a given velocity of movement the distance of movement (displacement from point of origin) is, of course, a direct function of the duration of the signal. Therefore, by increasing or decreasing the duration of the signal the extent of movement was increased or decreased. The distance of movement could be reliably varied in a step-wise manner of 18 microseconds of change in interaural delay for each increment of increase or decrease in signal duration.

Initially an attempt was made to secure thresholds by merely decreasing the distance of movement to the point where the movement was just perceptible. However, it was found that *Os* gave thresholds that were unreliable even after a considerable period of training. To increase reliability it was necessary to use a procedure which required a discrimination between the signal during movement and when it was stationary at the point of origin of movement. Thresholds were specified by decreasing the distance of movement in successive increments, and at each step both a moving and a stationary signal were presented. The order of signal presentation at each step, either moving or stationary was, of course, systematically randomized. The *Os* were instructed to respond by pressing a key after the cessation of each signal only if they were sure that it moved. At each step in the series the duration of both the moving and stationary signal was the same.

A counterbalanced design was used with at least ten threshold measurements (descending series) for the two signal noise bands at the different velocities of movement being obtained during a single experimental period for each location of movement. The measurements were checked during a second experimental period and an average taken of the threshold values during the two periods. With the exception of the two lowest velocities of movement, *Os* reported successive threshold values which were consistent within one to two increments of change in distance of movement (i.e., 18 to 36 microseconds of interaural delay). At the two lowest velocities the range of successive threshold values increased with variances of the order of two to four increments of change in distance of movement. This is explained by the fact that at the lowest velocities the discrimination between a moving and stationary signal near the displacement threshold level was exceedingly difficult.

Results

The results with five *Os* are plotted in Fig. 1. The curves in the lower part of Fig. 1 show the displacement thresholds for the two signal noise bands with movement from the median and lateral plane and with the six velocities of movement. The curves in the upper portion of Fig. 1 show the signal duration at the threshold level under the various conditions of movement. The velocity and displacement at the threshold level are expressed in microseconds of interaural delay and in terms of the approximate displacement angle. The displacement angle was estimated by the formula suggested by Woodworth (**5**) for computing direction angles for interaural time differences with a free field source of sound that is close to the head. Computations of direction angle based on this formula for a given interaural delay generally agree with subjective reports of perceived direction angle of a signal presented with earphones and with the same interaural delay. Therefore, the displacement angles given are probably a fairly good approximation.

With the exception of the highest velocities of movement, the data show that the 106–1200 cps signal noise band produces lower threshold values with movement from both the median and lateral planes than does the 2400–3400 cps noise band. This would be expected as the literature on localization has shown that low frequency

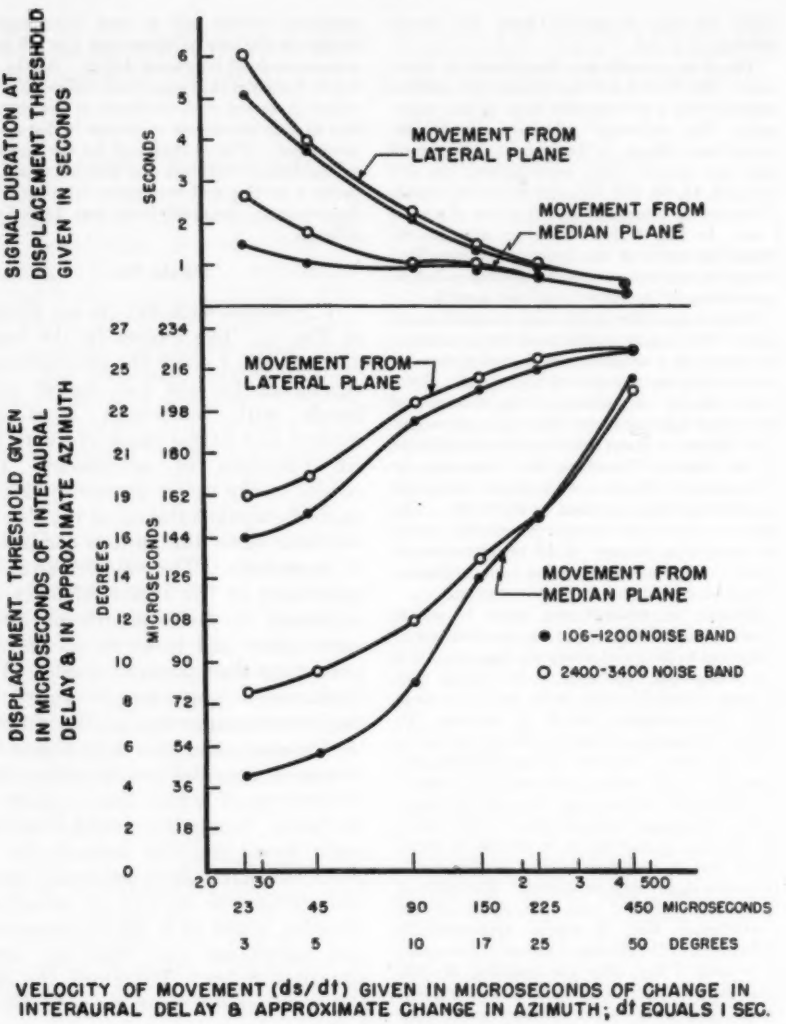


FIG. 1. Displacement thresholds at the various velocities of movement (lower section) and signal durations at the threshold level under the various conditions of movement (upper section).

sounds are localized more accurately than midrange frequencies. Also to be expected both on the basis of Pierce's results and previous localization studies is that the threshold values for movement from the median plane are lower than those with movement from the lateral plane.

The most interesting result is the relation between the velocity of movement and the displacement threshold. Under all the conditions

of movement the threshold level increases as the velocity increases. A further point for consideration is that the signal duration at the threshold level for the various velocities, plotted in the upper part of Fig. 1, increases as the threshold decreases. These two relationships indicate that the minimum movement perceptible is a function of two inversely related variables, duration and displacement of the signal.

Discussion

The results of this study appear to be in good agreement with the literature on localization. The major point for explanation is the relation between the velocity of movement and the displacement threshold level. This can be accounted for by making the assumption that the perception of auditory movement depends on a signal successively passing through at least two separately localized points. Further, it is reasonable to assume that in order for the minimum differential localization threshold to be obtained (i.e., minimum level that two signals are heard as being located separately in phenomenal space), a minimum signal duration would be required. Thereby, if the latter assumption is correct, when a signal duration is less than this minimum, a larger differential threshold would be obtained. On this basis, it would be expected that at a high velocity of movement a signal would have to travel further in order for movement to be heard. The reason for this is that the duration of the signal at any given point in phenomenal space is less than that required for the least differential localization.

With regard to the foregoing discussion it is interesting that studies of visual movement have shown the opposite relation between the velocity of movement and the displacement threshold. With visual movement the displacement threshold is low for high velocities and increases as the velocity decreases (1). This may indicate a basic difference between the mechanism of visual space perception and those involved in auditory space perception. This suggests that further study of auditory movement would prove of value in determining the processes involved in auditory perception.

The displacement threshold at the lowest velocity of movement reported here probably cannot be taken as the minimum extent of movement perceptible as it can be seen from the curves in Fig. 1 that a plateau has not been reached. The lowest threshold levels obtained are higher than those reported by Pierce in his study with the sound cage. Also, Klumpp (2), using an electrical delay line, has found that the minimum differential localization from the median plane of a low frequency noise band is of the order of 12 microseconds of interaural delay, or about 1° of azimuth. Apparently, lower velocities would have produced lower displacement thresholds, but these were not investigated as the discrimination between a moving and stationary signal at very slow rates of movement was exceedingly difficult. A procedure which would probably permit a more accurate measurement of the minimum perceptible movement would be to hold the distance of movement constant and vary the velocity of movement.

It is to be emphasized that the movement employed was simulated by a linear step-wise change in interaural time delay and was therefore not the same as real movement in azimuth. During real movement varying interaural intensity differences would also be present, and in the case of movement toward or away from the head, doppler effect would be present. However, as the movement simulated by this procedure is clearly perceived, it is reasonable to assume that the functional relationships obtained can be generalized to real azimuth movement.

Summary

The angular displacement threshold of auditory movement simulated by a linear step-wise change in interaural time delay was determined under various conditions. Variables investigated were: (*a*) signal noise bands of 106–1200

cps and 2400–3400 cps, (*b*) movement originating at the median and lateral planes, and (*c*) six different velocities of movement ranging from a perceptually slow to fast movement. Lower thresholds were obtained with the low frequency noise band, and for both noise bands lower thresholds were obtained for simulated movement from the median plane. The threshold is directly related to the velocity of movement; the threshold increases as the velocity of simulated movement increases.

References

1. Basler, A. Über das Sehen von Bewegungen. I. Die Wahrnehmung kleinster Bewegungen. *Arch. Ges. Physiol.*, 1906, **115**, 582–601.
2. Klumpp, R. G. Discriminability of interaural time differences. *J. acoust. Soc. Amer.*, 1953, **25**, 823.
3. Pierce, A. H. *Studies in space perception.* New York: Longmans, Green, 1901.
4. Wilcott, R. C., & Gales, R. S. Comparison of the masked thresholds of a simulated moving and stationary auditory signal. *J. exp. Psychol.*, 1954, **47**, 451–456.
5. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received April 17, 1954)